U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 2005

Tactical Communications Training Environment for Unmanned Aircraft System Operators

Kevin Sullivan, Susan R. Flaherty
Aptima, Inc.

December 2016

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Aptima, Inc.

Technical Review by

Dr. William Bickley, U.S. Army Research Institute
Dr. Thomas Rhett Graves, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: Department of the Army, US Army Research Institute, 6000 6th Street Building 1464/Mail Stop: 5610, Fort Belvoir, VA 22060-5610

FINAL DISPOSITION: Destroy his Research Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE:
This material is based upon work supported by the Army Research Institute for the Behavioral and Social Sciences (ARI) under Contract No. W911NF-14-C-0110. The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 12-15-2016	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 05-08-2014 to 03-16-2016

4. TITLE AND SUBTITLE Tactical Communications Training Environment for Unmanned Aircraft System Operators	**5a. CONTRACT NUMBER** W911NF-14-C-0110
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 644775
6. AUTHOR(S) Kevin Sullivan and Susan R. Flaherty	**5d. PROJECT NUMBER** A792
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Aptima, Inc. 12 Gil Street, Suite 1400 Woburn, MA 01801	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	**10. SPONSOR/MONITOR'S ACRONYM(S)** ARI **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** Research Report 2005

12. DISTRIBUTION/AVAILABILITY STATEMENT:
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Dr. Scott Graham

14. ABSTRACT
In aviation operations, there is an increased need for effective and efficient opportunities for UAS operators to learn critical communication and teamwork skills. The Night Vision Tactical Trainer - Shadow (NVTT-Shadow) was developed as a game-based desktop solution to train tactical communications skills for the UAS payload operator. It is intended as supplementary training for U.S. Army 15W Advanced Individual Training Soldiers (AIT) to learn and practice MUM-T tactical communications skills. NVTT-Shadow combines speech and text recognition, and natural language processing, with a human performance measurement platform to provide interactions, in a nominal mission environment, thereby enabling training of critical voice communication skills. Real-time audio feedback was provided to Soldiers throughout communications with virtual entities. Access to remediation and cumulative performance scores was available via after action reviews. An initial user assessment was conducted at Ft. Huachuca AZ with eight advanced individual training Soldiers and UAS course instructors. The usability testing demonstrated the feasibility of interactive gaming applied to MUM-T tactical communications. Ratings and comments from both students and instructors validated the need as well as mission context, game content, and game relevance.

15. SUBJECT TERMS
Manned-Unmanned Teaming, Communications Training, Unmanned Aircraft Systems operators, Simulation-based Training

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Scott E. Graham
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	Unlimited Unclassified	53	**19b. TELEPHONE NUMBER** 706-545-2362

Research Report 2005

Tactical Communications Training Environment for Unmanned Aircraft System Operators

Kevin Sullivan, Susan R. Flaherty

Aptima, Inc.

Fort Benning Research Unit

Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street Bldg. 1464
Fort Belvoir, Virginia 22060

December 2016

Army Project Number 633007A792	Personnel Performance and Training

ACKNOWLEDGEMENT

The authors extend their great appreciation to those who helped conduct the research and development described here: Dr. Martin Bink and Dr. Scott Graham, (ARI), Dr. Fredrich Diedrich, and all our contributing partners at Imprimis Inc., ASTi Inc, Aptima Inc., Kinex Inc., and Trideum Inc. The authors also thank our colleagues at Night Vision and Electronic Sensors Directorate, Modeling and Simulation Division, Ms. Susan Harkrider and Mr. Chris May. Finally, the authors thank U.S. Army 2-13th AV Regiment, Ft. Huachuca, AZ for their participation and support toward the advancement of this research.

TACTICAL COMMUNICATIONS TRAINING ENVIRONMENT FOR UNMANNED AIRCRAFT SYSTEM OPERATORS

EXECUTIVE SUMMARY

Research Requirement

U.S. Army Aviation is organically teaming rotary wing aircraft with Unmanned Aircraft Systems (UAS) as a force multiplier in combat operations. Clear and effective communications between the teamed airborne helicopter pilot and the UAS operator in the ground control station are critical for mission success. There is a documented need for effective and efficient opportunities for UAS operators to learn critical communication and teamwork skills. Consistent with the Army Learning Model, these opportunities should enable learning across institutional, operational, and self-development domains, and leverage new training capabilities that facilitate practice with associated feedback tools to guide learning.

Procedure

The Night Vision Tactical Trainer - Shadow (NVTT-Shadow) was developed as a game-based desktop solution to train tactical communications skills for the UAS payload operator (PO). The game is intended for U.S. Army 15W Advanced Individual Training Soldiers (AIT) to learn and practice Manned Unmanned Teaming (MUM-T) tactical communications skills. The NVTT-Shadow training game is composed of several Government Off-the-Shelf (GOTS) simulation components including 1) One Semi-Automated Forces (OneSAF), 2) AVSim Flight Model, 3) Night Vision Image Generator (NVIG), 4) Voice recognition server (ASTi Voisus), and 5) Aptima's Performance Measurement suite. NVTT-Shadow guides the trainee through a series of mission engagements using tactically relevant scenarios involving a variety of manned, unmanned, aerial, and ground-based assets.

The trainee is the only live player in the game and he must collaborate, from his web-based mock operating station, with various synthetic players via spoken natural language over simulated radio in order to execute the training missions successfully. The trainee is given feedback on the accuracy, completeness, order and timeliness of their communications across increasingly difficult scenarios. This report provides a description of the NVTT-Shadow system development, the integrated communications measures, and mission scoring and feedback.

Findings

An initial user assessment was conducted at Ft. Huachuca AZ with eight Soldiers and qualified UAS course instructors. While there were no statistically significant differences between pre and post ratings of MUM-T communication knowledge and skills, the usability testing demonstrated the feasibility of interactive gaming applied to MUM-T tactical communications. Acceptance from both populations validated mission context, game content, and game relevance. Limitations of the study included sample size and duration of gameplay.

Utilization and Dissemination of Findings:

The NVTT-Shadow system, along with a final project briefing and game software demonstration was given to Training Development and Battalion leadership at 2-13th AV Regiment, Ft. Huachuca, in March 2016. The results were also presented at the Interservice/Industry Training, Simulation, and Education Conference in Orlando, FL, December 2016. The system is being upgraded by the Communications-Electronics Research,

Development and Engineering Center’s Night Vision and Electronic Sensors Directorate to include enhancements recommended in the evaluation.

TACTICAL COMMUNICATIONS TRAINING ENVIRONMENT FOR UNMANNED AIRCRAFT SYSTEM OPERATORS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

TACTICAL COMMUNICATIONS TRAINING ENVIRONMENT FOR UNMANNED AIRCRAFT SYSTEM OPERATORS

Introduction

The role of Unmanned Aircraft Systems (UAS) is rapidly evolving from a traditional intelligence and surveillance role to a more active participant as a scout-reconnaissance asset that can designate and destroy targets. Accordingly, Manned-Unmanned Teaming (MUM-T) is emerging as a critical element of aviation operations in which UAS are teamed with rotary wing assets to conduct operational missions. A primary training challenge is that UAS operators traditionally learn few of the scout-reconnaissance skills appropriate to MUM-T at the schoolhouse (Stewart, Bink, Barker, Tremlett & Price, 2011). Teaming between manned and unmanned air and ground assets will become increasingly prevalent, necessitating training strategies and capabilities that will build the required skills for mission effectiveness. There is a need, therefore, to provide earlier opportunities for UAS operators to learn critical communication and teamwork skills. Consistent with the Army Learning Model (ALM; Department of the Army, 2011a), the training should enable learning across institutional, operational, and self-development domains, leverage technology-based training as appropriate, and provide frequent feedback to guide learning.

Following principles outlined in *The U.S. Army Training Concept 2012-2020* (Department of the Army, 2011b), a skills-adaptive, game-based desktop solution was developed to train tactical communications skills for the UAS payload operator. The game is intended for U.S. Army Unmanned Aircraft Systems Operator (15W) Advanced Individual Training Soldiers (AIT) who have completed a minimum of Phase 1 Common Core Aviation ground school. In collaboration with ARI and U.S. Army Night Vision and Electronic Sensors Directorate (NVESD), the Night Vision Tactical Trainer - Shadow (NVTT-Shadow), a tactical communications training game, was developed. NVTT-Shadow provides nominal MUM-T mission scenarios (see United States Army Aviation Center of Excellence, 2014) to learn and practice MUM-T tactical communications skills in a single player domain.

This report describes the development of the NVTT-Shadow training system, including the development of an automated voice communication scoring system and MUM-T training scenarios. The report also includes the results of an initial usability assessment with Soldiers in in initial UAS training and UAS course instructors.

NVTT-Shadow System Development

Game Overview

The NVTT-Shadow training environment is composed of several Government Off-The-Shelf (GOTS) simulation components depicted in Figure 1, which guides the trainee through a series of increasingly complex engagements using tactically relevant scenarios involving a variety of manned, unmanned, aerial, and ground-based assets. The trainee is the only live player in the game and he must collaborate, from his web-based simulated control station, with various

synthetic players via spoken natural language over simulated radio in order to execute the training missions successfully. Multiple single players may be run simultaneously. Synthetic entities are modeled in two complementary layers -- One Semi-Automated Forces (OneSAF) provides basic background behaviors for entities, while NVTT-Shadow provides higher-level models that control entity actions based on intent extracted from the trainee's spoken natural dialog with game entities. Dialog structure is modeled based on Army standards for communication and verbal protocols (U.S. Army Aviation, 2014).

The training game comprises the following key elements: (a) UAS simulation platform, (b) a commercial speech recognition system, (c) natural language parsing system, (d) human performance measurement system, (e) communications performance measures, and (f) UAS mission scenarios. Technical challenges addressed in this project include: management of the dialog state vs. scenario context between the trainee and synthetic entities during speech recognition; design and implementation of communications measures based on pattern and vocabulary matching synchronized to simulation events; and the development of context sensitive feedback to guide the trainee based on measurement results.

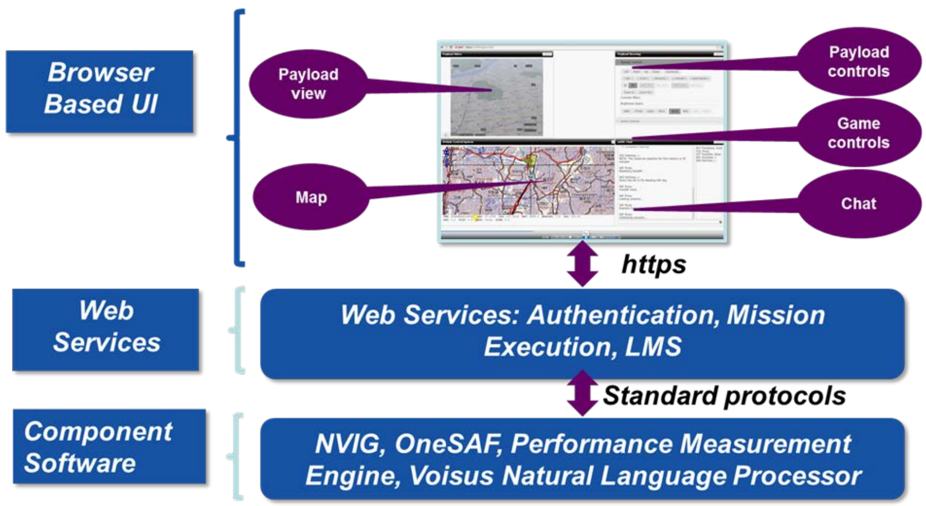

Figure 1. NVTT-Shadow Architecture

NVTT-Shadow was designed to augment the current UAS Operator Program of Instruction (POI) by focusing on MUM-T related critical tasks with specific focus on accurate and timely tactical communications. Gameplay is intended to supplement and reinforce instruction in existing POI modules. For example, NVTT-Shadow gameplay could be used to enhance Soldier knowledge retention while awaiting to participate in practical flight line skill exercises. The game aims to develop tactical communication skills proficiency of the payload operator through the execution and accomplishment of mission objectives in routine UAS missions (e.g., route reconnaissance, convoy security).

Coordination with external agencies and manned aircraft scout/attack teams are emphasized. Game content encompasses MUM-T related critical tasks outlined in the UAS

Commander’s Guide and Aircrew Training Manual (TC 3-04.61; Department of Army, 2014) and are included as Table 1 and 2. More complex mission tasks involving UAS gunnery and fire training are based on skills outlined in Combat Aviation Gunnery (USAACE, 2014).

The NVTT-Shadow game progresses the Soldier through a series of UAS scout mission scenarios involving various assets (e.g., manned, unmanned, aerial, and ground-based). The Soldier, while seated at a laptop simulated Ground Control Station (GCS), interacts with virtual team entities via ad hoc communication over a simulated radio. NVTT-Shadow is a web-based system integrating natural language processing components (i.e., speech recognition, speech-to-text, text-to-speech, and language recognition), performance measurement and feedback components into a (OneSAF) flight simulator platform.

Table 1
RQ-7 Shadow Aircrew Training Program Requirements

Task #	Title
1110	Track a Static Target
1115	Track a Moving Target
1120	Perform Aerial Reconnaissance
1125	Call for and Adjust Indirect Fire
2054	Perform Target Hand Over to an Attack Helicopter
2092	Transmit a Tactical Report
2474	Designate for a Laser Guided Missile

Table 2
MUM-T Tasks Outlined in Combat Aviation Gunnery Represented in Game Missions

UAS Common Tasks
Target Handover Using the Laser Target Marker and Laser Designator
Request and Adjust Indirect Fire
Request Close Combat Attack (AH-64 cannon & rockets)
Request and Designate Remote Hellfire Missile
Request Close Air Support (Bomb on Coordinate)

Game Components

NVTT-Shadow operates as a distributed, web-based game supporting multiple, simultaneous players executing unique missions. The minimum (i.e., single player) hardware configuration is composed of two rack-mounted software servers displaying on a 22-in 1920x1080dpi monitor with keyboard and joystick. A laptop can be used by a game proctor/ instructor as a duplicate of the Soldier’s game display. The ability to monitor correct transcription of speech-to-text is also accessible via the voice client panel. Game software comprises of the following five integrated simulation services that function across a Distributed Interactive Simulation (DIS) interface to generate the mission training environment: (a) OneSAF,

(b) AVSim Flight Model, (c) Night Vision Image Generator (NVIG), (d) Voice recognition server (ASTi Voisus), and (e) Aptima's Performance Measurement suite.

NVTT-Shadow utilizes a geographically diverse terrain database depicting terrain features including mountains, rivers, valleys, and coastal regions. Correlated satellite imagery overlays the wireframe terrain world, such that cultural features (e.g., bridges, roads) are functionally represented. The presence of simulated ground vehicles and dismounted Soldiers is defined within OneSAF and subsequently overlaid onto the terrain database. A single terrain database is utilized for all game missions. Figure 2 shows the NVTT-Shadow user interface with (clockwise from upper left) a voice client, terrain view, sensor and laser control panels, chat, and mission map windows.

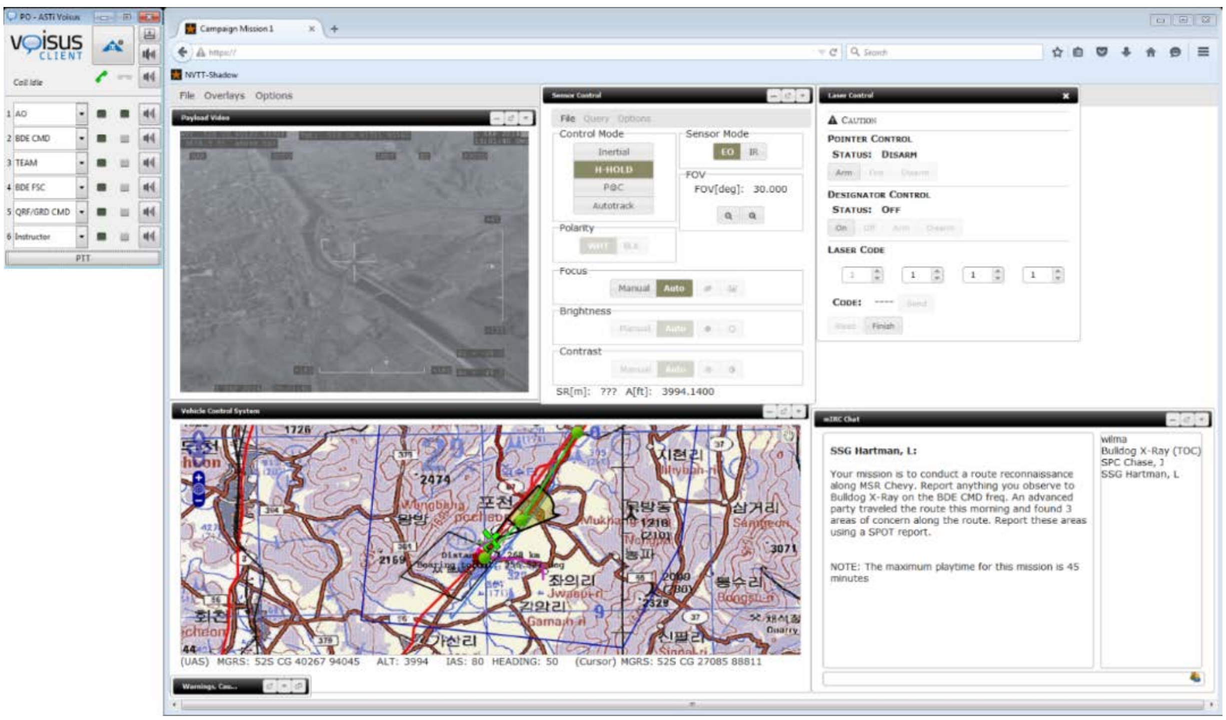

Figure 2. NVTT-Shadow User Interface

Game Missions

Game scenarios are divided into Training missions and Campaign missions. Training missions exercise progressively more difficult skills in increasingly complex situations and are unscored. These missions serve as a tutorial for understanding command and control of the aircraft and payload sensors, as well as providing training on all tactical communications required for successful execution of the Campaign missions. Soldier interaction is prompted and guided by a virtual instructor toward achieving the learning objective. Training missions are indexed by the primary learning objective/tactical report contained within. By contrast, Campaign missions are designed for Soldiers to exercise free play without guidance from a virtual instructor. Campaign missions are scored for timing and communications accuracy. Each Campaign mission begins with a briefing outlining the high-level mission objective (e.g., area

reconnaissance) and critical actions required for success (e.g., “locate hostile threats and conduct target handover to Apache 11 using laser target marker”). Feedback and measurement of mission objectives as well as tactical communications are displayed upon mission completion.

Training missions provide a way for payload operators to familiarize themselves with the NVTT-Shadow user interface components by presenting a series of unscored, highly constrained missions. These missions are meant to familiarize the payload operator to NVTT-Shadow peculiarities, e.g., differences between operational Ground Control Stations and the NVTT-Shadow mock ground control stations. Training missions also provide a means to reinforce knowledge of the structure and cadence of tactical messaging.

Throughout the training missions, a constructively modeled instructor guides the payload operator through locating targets as well as the specifics of each verbal exchange, helping the payload operator to "think before he/she talks" and reinforcing his knowledge of the content and order of each exchange. The set of training missions offered by NVTT-Shadow begins with relatively simple verbal exchanges, e.g., SPOT reports. As the payload operator progresses through the missions, more complex exchanges with different combat partners are taught including battle damage assessments, target handovers to ground and air platforms, remote hellfire engagements, and calls/adjustments for indirect fire.

To complement the highly constrained and directed training missions, NVTT-Shadow provides a series of campaign missions that are almost entirely unconstrained. The constraints that exist for these missions come from a mission operations order and flight brief, which are presented to the payload operator just prior to the mission. Once the mission starts, it is up to the payload operator to execute based on this information, which is also available in-game via a kneeboard-like interface. In training missions, the instructor might provide a specific grid location for a target or correction if the payload operator, for example, asks the aircraft operator to fly outside of an ATC control measure. In campaign missions, the payload operator may get direction from the tactical operations center (TOC) or other entity, which may be less precise with respect to possible target locations and he may learn of mistakes like violating ATC control measures via radio transmission from the TOC or other non-human player. In training missions, the instructor will, for example, tell the payload operator to send a SPOT report or Battle Damage Assessment (BDA) to “Bulldog-Xray on the BDE CMD net.” In these less directed missions, attempting to contact another game entity on the wrong network will usually elicit a "check frequency" response or simply silence in response. It is up to the payload operator to realize the mistake and correct it.

Speech Data Processing

The Soldier uses a virtual multi-channel radio panel to communicate with game entities to affect aircraft maneuvers, weapon fires, support calls, and reconnaissance reports. Entities respond with real-time feedback based on Soldier speech and context. The game allows the Soldier to select which channels to transmit and receive on, ensuring that he/she remains aware of the automated entity with whom the Soldier is communicating. The system utilizes automatic speech recognition (ASR) tuned with application specific data followed by a natural language processing (NLP) stage that extracts relevant meaning from the text. Multiple potential

utterances can be collapsed into a single “extracted” meaning. For example, a trainee utterance of “Fly east” will have the same extracted meaning, as “Fly heading 090.” The extracted meaning, based on both the content of the trainee transmission and the current game context within the mission, is used for real-time feedback as well as later performance scoring. A text to speech (TTS) library generates realistic transmissions for real-time vocal responses to the Soldier via phraseology for SPOT and Battle Damage Assessment (BDA) reports, remote hellfire engagements, and other radio interactions

The meaning of the Soldier communications extracted by the natural language processor are passed to the game controller software to be validated and trigger in-game actions. The current context of the mission and state of the simulation are tracked in the form of a state machine. The game controller also maintains a state machine for each synthetic entity, and both state machines reside in a centralized context. Both voice system and the game controller can update the context, in response to voice commands or in-game events, respectively. For example, upon the Soldier transmitting to the virtual Apache pilot, “Remote hellfire,” the voice system updates the virtual Apache pilot’s state to reflect that the trainee is requesting remote hellfire air support. The game controller confirms that the Apache is able to perform the requested actions and updates its context, triggering the voice system to respond affirmatively using text to speech.

Context-based error checking is employed so that a game entity can provide real-time feedback. For example, if a Soldier calls for a remote hellfire engagement from the Apache pilot and provides an invalid grid location, the Apache pilot will respond, “Say again grid.” This tailored feedback takes into account the specific type of information expected based on the current game context and provides feedback to help the Soldier improve his communications.

Using language processing techniques, key information from Soldier transmissions is organized into utterances for performance assessment. Speech events and audio are shared over the DIS network to support the communications performance measurement system. Voisus is a suite of communications and sound products that provides the two-way communications pipeline by incorporating commercial communications products for speech recognition and for speech generation in to the NVTT-Shadow environment to enable accurate speech to text and text to speech functionality. Figure 3 illustrates the speech to text process.

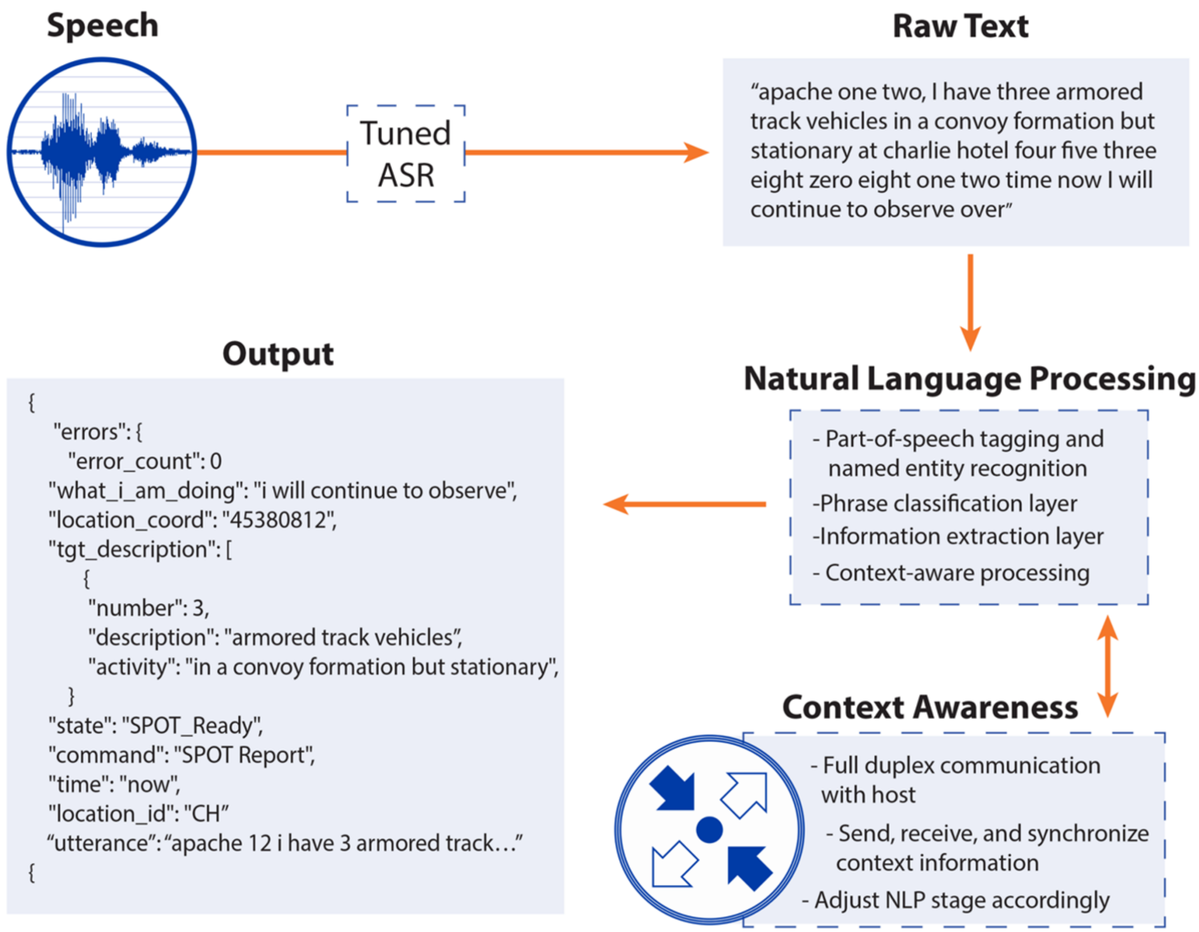

Figure 3. Speech to Text Processing

A transactional model for synchronizing context state machines is employed to manage the critical synchronization of the interdependent voice and game contexts. Any state change made by one system is confirmed by the second system prior to the second system changing. This mechanism allows Soldiers to issue commands, transmit information, receive relevant and constructive real-time vocal feedback, and visually confirm events occur in the game using only voice commands. A detailed discussion of the challenges and solutions employed in NVTT-Shadow regarding coordinating and synchronizing communication between trainees and synthetic entities is described in Berglie & Gallogly (2016).

Communication Measures

A critical component of the training game is the measurement and feedback of Soldiers' performance of accurate and timely tactical communications. Soldiers are assessed and scored on transmission and content of their radio transmissions across four primary dimensions:

1. **Accuracy**: Did the trainee accurately describe and report the event in the scenario? Trainee utterances must match one of a set of predefined possible lexical formulations for the event adhering to Army scout reconnaissance protocol (USAACE, August

2014). Moreover, specificity counts: for example, “red truck” is always preferred to simply “truck.” Distinctions such as these are reflected in the accuracy score.

2. **Completeness**: Did the trainee report all the required information for the event? Utterances are parsed into fields of required information with respect to communication type. For example, for a SPOT report, fields include number, description, activity, location, time, and “what I’m doing.” Completeness is computed as the percentage of slots filled by the trainee.

3. **Order**: Did the order in which a trainee reported the event match protocol? Most communication types must follow a structured format where the order of slots of information is prescribed. Order is computed as the distance in terms of “edits” (re-arrangement of a pair of slots) from the prescribed order.

4. **Timeliness**: Did the trainee report the event in a timely manner according to protocol? Timeliness is defined as the speed that a communication is formulated and transmitted relative to event occurrence in the scenario.

Measurement of these dimensions takes into account the content and form of Soldier utterances, as well as contextual information from the game environment to create a context-aware measurement system. For example, accuracy of a description in a SPOT report is relative to a known event in the scenario, and timeliness is measured with respect to the event onset in simulation runtime. More detailed discussions of types of measurements (communications vs. behavioral), as well as the algorithms for pattern matching and weighted scoring of speech utterances to expected communications protocol is included in the Appendices D and E.

Context aware (i.e. game situation aware) performance measurement affords a semantic level of measurement of situations where virtual entities are participants. For example, the trainee may view a group of persons firing weapons at a building. The natural language parsing engine aligns the game situation with a dictionary of utterances to supplement adherence scoring. Thus, beyond just providing a description of, “persons firing weapons” in the SPOT report, the trainee may also provide a description of “terrorist attack” and still receive a partial grade for still conveying the battlefield threat to the commander. The extension to context awareness allows for domain specific speech to be scored more favorably than common jargon.

Assessment along the measurement dimensions is applied to the raw values to bin the Soldier’s score as “expert,” “average,” or “poor” performance. Performance assessment thresholds were established from input by Army Aviation Subject Matter Experts (SMEs). Measurement flexibility is supported by performance thresholds that can be tuned by an instructor to meet alternate goals.

Mission Scoring and Feedback

A player’s primary pre- and post-mission interaction with the game is through the game’s scorecard (Figure 4). The trainee uses the scorecard to choose which mission to play. After completing a mission, a trainee views the scorecard to receive feedback on their performance.

The scorecard contains the following elements: Campaign Mission Map, Training Mission Overview, Mission After Action Review (AAR), and Mission Selection.

The Campaign Mission Map was designed to orient the student to their overall progression through the game. The overarching campaign is composed of missions with a unifying objective. At a glance, the student is able to ascertain their progress through the game as well as individual mission accomplishment. Prior to mission play, a student may select a mission to receive either a mission briefing or their After Action Review (AAR) if previously played.

Figure 4. NVTT-Shadow Mission Map and Scorecard

The game interface (Figure 4.) is divided into three sections of information. First, the Banner area displays basic user information including their name, game rank, total score, as well as number of missions played. Second, the Mission Map displays all of the missions in the game. Hovering and clicking reveal greater mission detail. Third, the UAS Dashboard displays selected mission details including training objectives, previous play, and averaged scores.

The Training Mission Overview provides the mission essential information. Players are given a list of training missions along with the corresponding Training Objectives. A Soldier may either follow the guided order of training or choose to focus on selected training by need.

Soldier feedback in the Mission AAR was designed based on the recommendations for formative feedback by Shute (2008). The guiding philosophy outlines focusing on the task, limiting the feedback to only specific areas with cues for student improvement. The feedback is both situation and context-specific. Once Soldiers select a specific communications report, they

were provided four types of feedback: (a) an assessment of the communications dimension, (b) context-specific narrative feedback, (c) access to his voice recording, and (d) an exemplar recording from an expert. By providing different levels of feedback, the student is able to receive the associated cue for the task being assessed. For example, the student may listen to the audio samples to hear the difference in cadences between himself and the expert. Additionally, he is able to view accomplishment (solid green), or deficiency (red) of the mission objective by viewing the overlay.

Figure 5. NVTT-Shadow Scorecard Detail and Performance Feedback

Detailed Soldier performance vs. game objectives and communications measures is presented in four sections (Figure 5).

1. Training Objectives are shown with both the number of attempts at that objective as well as their assessment.
2. Situations described in the trainee's reports are shown for each.
3. The assessments for each of the communications dimensions is shown along with narrative feedback
4. The trainee may either listen to their speech for the report or that of an expert giving the report.

In addition to the Player Dashboard, an Instructor Dashboard was developed to view the aggregate performance of multiple Soldiers. Design features include the ability to sort data by Campaign Mission, Training Objective(s), or Student(s). Sorted data is presented in graphical format to support any remediation deemed necessary by the instructor.

Usability Assessment

Participants

Four 15W UAS AIT Soldiers and four UAS instructors participated in the evaluation. AIT Soldiers included two Shadow operators and two Gray Eagle operators enrolled in their 15th and 32nd week of instruction, respectively. Two of the four participating instructors were dual-qualified on Shadow and Gray Eagle UAS platforms, with the remaining two qualified on Shadow UAS, exclusively. Demographic surveys indicated most of the Soldiers had experience playing video games, as they averaged playing 3.6 days a week and 1.7 hours a day.

Procedure

The assessment was conducted at Ft. Huachuca, AZ. Two participants were scheduled at a time for a two hour session of self-paced gameplay, including approximately 30 minutes of tutorial training missions for game familiarization. Prior to the training session, the Soldiers completed demographic and knowledge assessment surveys. The Soldiers rated their knowledge of nominal MUM-T tactical communications tasks (e.g., 9-line report, 5-line report, SPOT report). There was minimal proctor intervention. While one Soldier progressed through training and subsequent campaign mission, the second Soldier completed a demographic survey and pretest knowledge assessment on tactical communications protocol.

Soldiers were introduced to joystick payload functions and method of gameplay through a series of short training missions prior to the scored Campaign mission gameplay. Appropriate communications protocols for specified mission context were outlined in simple mission instances via interaction with an instructor avatar. Emphasis on appropriate selection of radio channel and sequence of information elements within a tactical communications report was given to advance the student through each learning objective. Feedback on correct transmission or need for improvement was audibly given by the instructor avatar in real-time. Training missions differed from Campaign missions in that they were ungraded and advancement was predicated on successful completion. Training missions required 5-15 minutes of gameplay depending on Soldier’s skill level.

The Soldiers then conducted two campaign missions. Mission success was predicated on accomplishing mission objectives in a timely and accurate manner with the appropriate tactical communications. For the purpose of this assessment, tactical communications included a SPOT report, Call for Indirect Fire, and Battle Damage Assessment. Campaign mission duration was approximately 30 minutes.

As Soldiers progressed through the Campaign missions, they received formative performance feedback. The feedback not only provided corrective information but also presented Soldiers with exemplar audio reports for learning remediation. As described above, Soldiers were scored on the transmission and content of tactical communications across four primary dimensions: Accuracy, Completeness, Order, and Timeliness.

Scoring took into account the content and form of the Soldier utterances, as well as contextual information from the simulation environment. For example, accuracy of a description in a SPOT report is relative to a known entity or event in the scenario captured from the DIS data stream, and timeliness is measured with respect to the event onset in simulation runtime.

After concluding the game session, the Soldiers again rated their knowledge related to tactical communications. Next, a user feedback questionnaire was administered to collect subjective ratings on: (a) Game Effectiveness and (b) Perceived Value of Game Practice to Specified Communication Reports. In addition to ratings, audio files were collected for post-assessment natural language processing refinement. Lastly, Soldier comments related to game preferences (i.e., likes, dislikes) were documented.

Results

There were no statistically significant differences between pre and post ratings of MUM-T communication knowledge and skills. This was most likely due to the small sample size (N = 8) and limited gameplay.

Knowledge Ratings. Soldiers rated their knowledge of nominal tactical communications reports required in MUM-T operations before and after exposure to NVTT-Shadow gameplay. Ratings were collected on a 4-point scale (1 = No Knowledge; 4 = Very Knowledgeable). Although no statistical significance was found between pre and post gameplay means (Figure 1), it is possible that longer gameplay would lead to higher post game playing ratings.

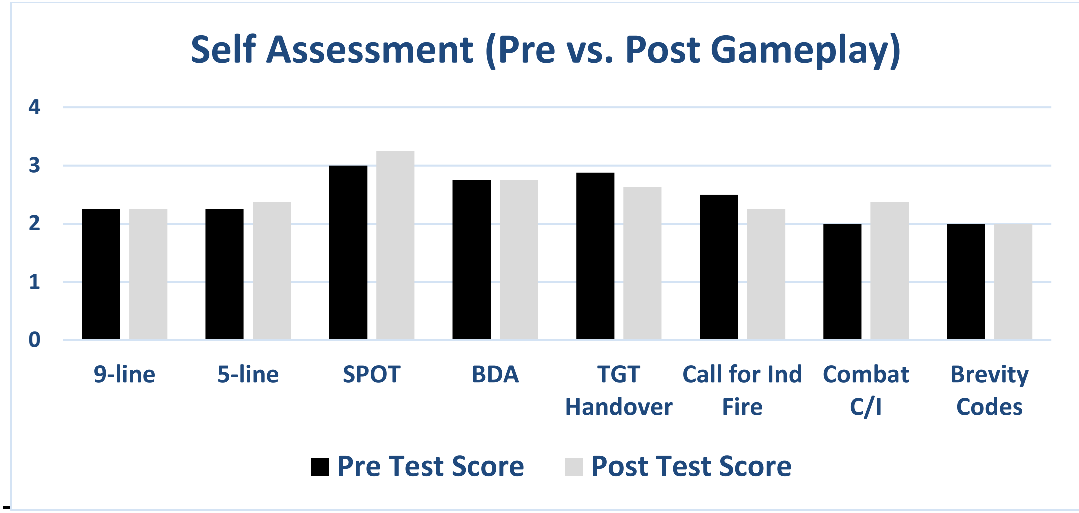

Figure 6. Knowledge Self-Rating Pre and Post Gameplay

Game Effectiveness. Soldiers rated game elements contributing to an overall profile of game effectiveness. Ratings were collected on a 4-point scale (1 = Major Deficiencies (Undesirable or Ineffective); 4 = Exceptional (Highly Desirable or Very Effective). A summary of mean ratings is presented below (Figure 7). It should be noted that ratings were closely

distributed, such that the standard deviation of each of the ratings was less than 1. Of the ten elements rated, six elements received an average rating of 3 or greater (green bars). Ratings of greater than 3 were deemed without deficiency and desirable to highly desirable. Game graphics received the lowest average rating, which was expanded upon in Soldier comments related to window obscuration and absence of hotkey mapping.

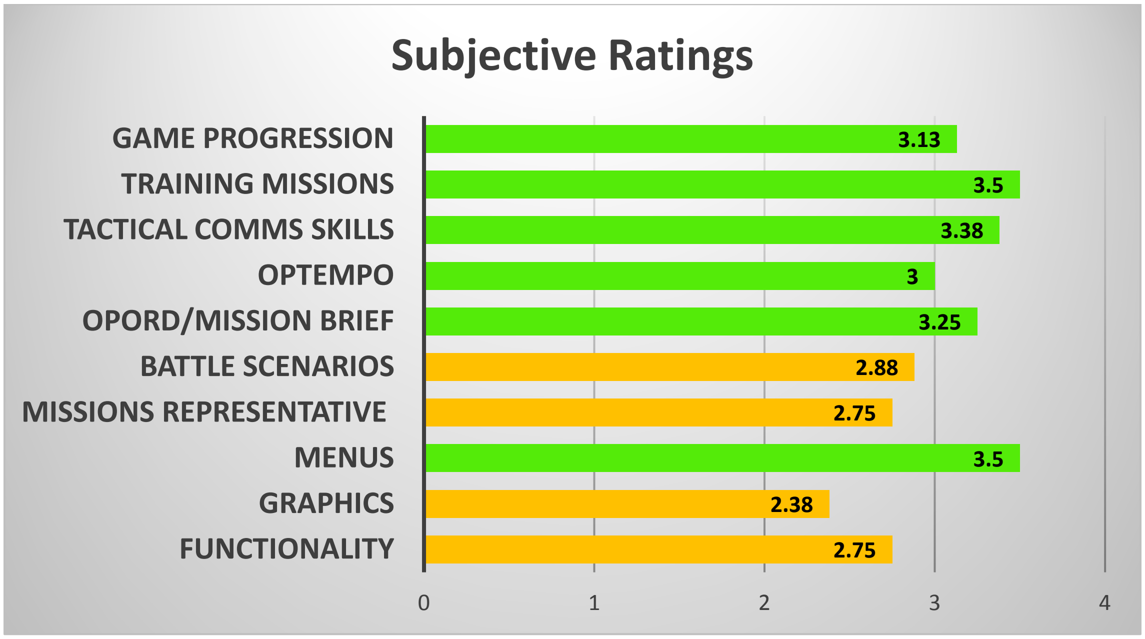

Figure 7. Soldier Ratings of Game Elements

Soldier Comments. The written responses identified areas for game improvement, desired additional features, and provided general user opinions about potential training value.

Joystick controllability was frequently cited as needing improvement. Difficulties in controllability of the payload due to unexpected or variant behaviors were the underlying reason given. Voice recognition performance was also identified as an area for improvement. As previously stated, one of the purposes of the usability testing was to collect additional voice data for further natural language processing refinement. Lastly, software failures to generate scorecards during some missions were documented.

Several Soldiers stated that payload autotrack performance should be enhanced to match the current aircraft capabilities. Also, the labelling of hotkeys and access to joystick mapping during the game was specified. On display technology, Soldiers stated that better contrast of map overlays would enhance ability to see borders and boundaries. Lastly, the possibility for a demonstration mode was mentioned in order for users to understand what is acceptable within gameplay.

Overall acceptance, relevance and perceived benefit of the game received excellent marks from course instructors qualified in both Shadow and GE platforms. Game instructions were found to be “clear and to the point.” Due the missions presented, slight favor was shown for

inserting gameplay into the simulation phase of Gray Eagle versus Shadow instruction. One instructor stated that he foresaw “astronomical benefit” to the solider trainee from playing the game over the current methods of voice communications training. An AIT Soldier affirmed, “It’s more interactive than what we are currently doing.” Gameplay was favored for overcoming the current challenge of the required 1:1 student/instructor contact when refresher or remediation training is needed in tactical communications skill training. Additional comments showed that AIT Soldiers believed the game was “very useful” and the “correct technology.” All eight Soldiers described the positive learning potential for MUM-T tactical communications practiced in a game setting.

Discussion and Recommendations

The primary intent of this Rapid Innovation Funding effort was to demonstrate an interactive medium-fidelity training environment for UAS operators to exercise MUM-T communications, and secondly to enhance trainee learning by providing a richer and more engaging After Action Review (AAR) on their communication performance. Use of the training environment should allow the course instructors to adapt to operator performance both in real time, by enabling the synthetic entities tailored to the skill level of the operator, and over the course of the curriculum, by providing detailed information on past course level performance which will allow for more data-driven selection of subsequent training material.

Developing game entities that could “understand and generate natural language and responsive behaviors relevant to the UAS sensor payload operator training scenarios” presented multiple technical challenges. Demonstrations and usability testing results pointed to needed improvements to refine and harden the system to be more flexible in accepting variant speech patterns and pronunciations. Trainee frustrations did occur, when trainee speech was not recognized by the game entities. Root causes for typical recognition gaps were found to be due to pronunciation variances across the trainee population. As a secondary effect, failures in speech recognition occasionally produced issues of dialog synchronization with the game controller. These issues and possible mitigations, among others, are discussed in detail in Appendix E. An additional stakeholder recommendation was that NVTT-Shadow campaign missions be tuned to more specifically align with the needs of the schoolhouse curriculum.

Secondary, but not insignificant was the challenge of determining what was actually a correct and acceptable natural language response. Again, as discussed in other sections, natural language provides wide variances in acceptable methods of communication. Improvements in speech recognition can be made via continued logging of student voice logs for tuning the speech recognition and communications measures thereby increasing the coverage of the breadth of variance in speech recognized. Expansion of the library of acceptable utterance patterns can come from both instructor input and by logging exemplar play by instructors to inform both the communication measures and speech recognition.

Future Development Efforts

The NVTT-Shadow simulation implemented scripted behaviors of artificial entities within the game (e.g. for enemy forces, vehicles, etc.) that occur at set times and locations. These

scripted behaviors do not always match up with the trainee game play actions. The trainee performing the reconnaissance flight might not maneuver his Shadow UAS to the location where and when the scripted behavior is occurring. The mismatch in trainee action and scripted actions can lead to a loss of a training opportunity. Significant improvements in the usability and flexibility of mission scenarios would likely occur if the system dynamically coordinated simulation scripted activities with actual trainee behaviors. Related to the limitations stemming from scripted entity behavior is the challenge in the current implementation of pausing, restarting or saving game sessions in the same entity and trainee state which is not currently supported. Saving state and supporting restart of sessions could lead to significant efficiencies in exposing trainees to training opportunities without having to play through whole games.

In general, the usability testing conducted with the 15W instructors and 15W AIT Soldiers demonstrated the feasibility of interactive gaming applied to MUM-T tactical communications. Acceptance from both populations validated mission context, game content, and game relevance. Clearly these observations are limited by the small sample size and duration of gameplay, but the observations and comments are encouraging. This research overcame many technical challenges. The research further demonstrated a training system that, while not perfect, meets an important training need and demonstrated the future of automated speech recognition in a tactical communications training environment.

References

Berglie, S. T. & Gallogly, J. J. (2016). Collaboration between human and nonhuman players in night vision tactical trainer-shadow. *Proceedings of SPIE 9848, Modeling and Simulation for Defense Systems and Applications XI, 984803,* Orlando, FL.

Department of the Army (2011a). The U.S. Army Learning Concept for 2015. (TRADOC PAM 525-8-2). Fort Monroe, VA: Author.

Department of the Army (2011b). *The U.S. Army Training Concept 2012-2020*. (TRADOC PAM 525-8-3). Fort Monroe, VA: Author

Department of the Army (2014). *Combat Aviation Gunnery*. (TC 3-04.45). Fort Rucker, AL: Author.

Flaherty, S. R., & Bink, M. L. (2015). Training manned-unmanned teaming skills in Army Aviation. *Proceedings of International Symposium on Aviation Psychology,* Dayton, OH: Wright State University.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153 - 189.

Stewart, J. E., Bink, M. L., Barker, W. C., Tremlett, M. L., & Price, D. (2011). *Training needs for RQ-7B unmanned aircraft system operators in the scout-reconnaissance role.* (ARI Research Report 1940). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADB367652).

Sultan, M. A., Bethard, S., & Sumner, T. (2014). Back to Basics for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidence. *Transactions of the Association for Computational Linguistics*, *2*.

United States Army Aviation Center of Excellence (USAACE). (August 2014). MUM-T Handbook: Leveraging Aviation Manned Unmanned Teaming (Draft). Fort Rucker, AL.

Acronyms

15W	Military Occupational Specialty for Unmanned Aircraft Systems Operator
AAR	After Action Review
AIT	Advanced Individual Training
AO	Aircraft Operator
APOE	Air Point of Entry
ARI	U.S. Army Research Institute for the Behavioral and Social Sciences
ASA	A-Measure Server Application
ASR	Automatic Speech Recognition
BDA	Battle Damage Assessment
BDE CMD	Brigade Command
CADRG	Compressed ARC Digitized Raster Graphics
GCS	Ground Control Station
GOTS	Government Off the Shelf
HPML	Human Performance Markup Language
IED	Improvised Explosive Device
LMS	Learning Management System
MOS	Military Occupational Specialty
MGRS	Military Grid Reference System
mIRC	An Internet Relay Chat service
MUM-T	Manned Unmanned Teaming
NLP	Natural Language Processing
NLU	Natural Language Understanding
NVESD	Night Vision Electronic Sensor Directorate
NVIG	Night Vision Image Generator
NVTT-Shadow	Night Vision Tactical Trainer - Shadow
OneSAF	One Semi-Automated Forces
POI	Program of Instruction
SPOT	SPOTREP or SPOT Report
TOC	Tactical Operations Center
TTS	Text to Speech
UAS	Unmanned Aircraft System
UAV	Unmanned Aerial Vehicle

APPENDIX A

SOFTWARE COMPONENTS

Architecture Overview

This effort extended a prototype system to create a more realistic training environment that allows operators to interact with more sophisticated synthetic entities using voice communications, and which includes an embedded performance assessment capability to guide learning. The training environment is web-based and enables trainees to interact with the trainer via a joystick and headset.

The result enables a more realistic training experience for UAS operators using the full communications spectrum, as well as presenting a richer AAR. These efforts will ultimately allow for the trainer to adapt to operator performance both in real time, by enabling the synthetic entities to consider operator performance when determining their next behavior or response, and over the course of the curriculum, by providing instructors with detailed past performance information allowing for more data-driven selection of subsequent training material.

The NVTT-Shadow software has been architected for both scalability and modularity. The component interfaces have been architected using Representational State Transfer (REST) principles for creating scalable web services. The software components have also been architected for scalability, where possible. The NVTT-Shadow simulator uses proxies to interface with components. The Measurement Platform provides a single interface that allow for resources, such as PM Engine, to cross domains. These approaches combine to create a distributed architecture that will scale natively.

The system architecture is shown in Figure A-1. The diagram is color-coded to indicate major components:

- Voisus – Green
- NVTT-Shadow Simulator – Brown
- OneSAF – Purple
- Measurement Platform – Orange

All components share the DIS Network indicated in blue.

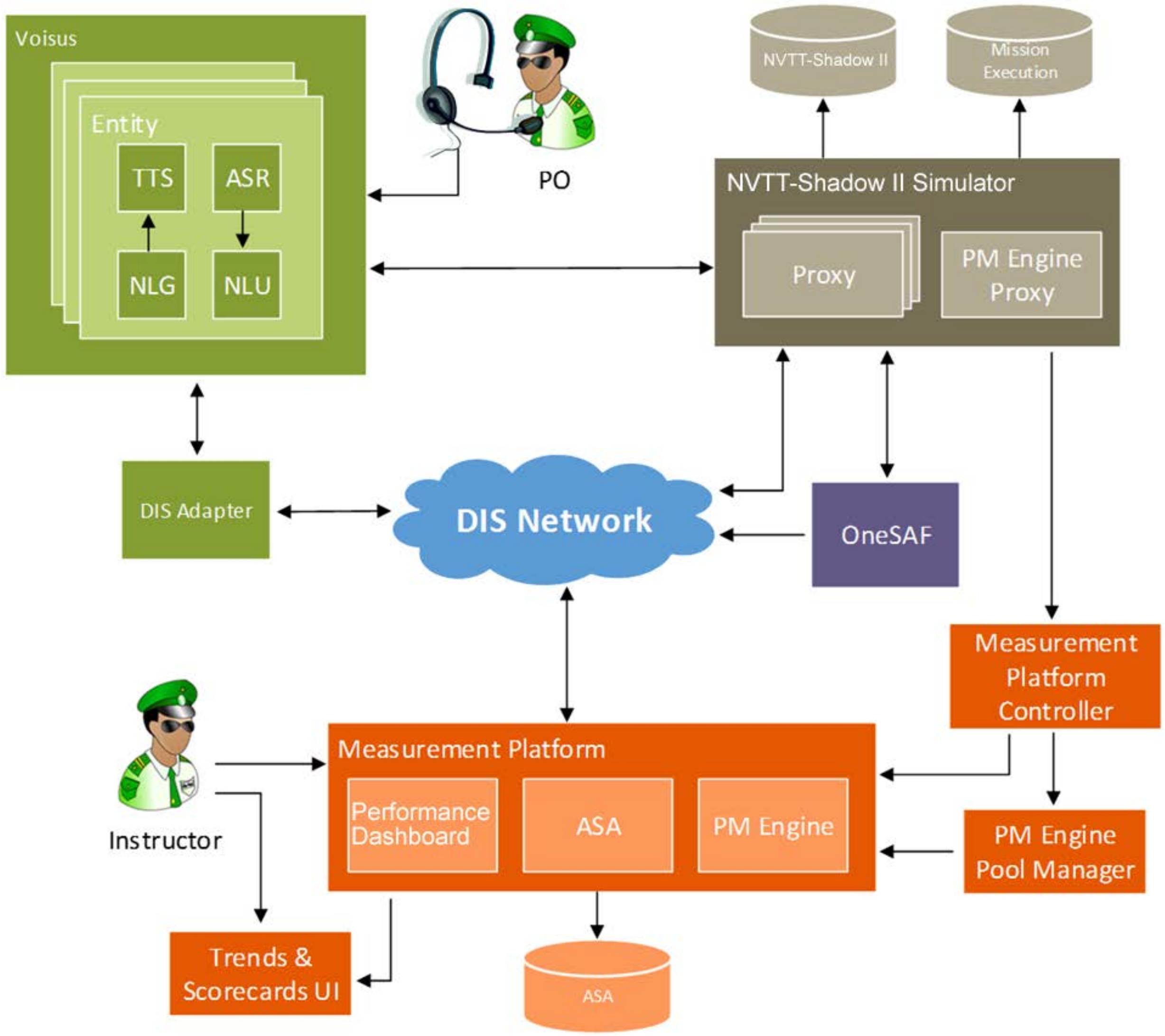

Figure A-1. System Architecture

The ***Voisus*** component consists of representations for each synthetic entity type and supports speech to text and text to speech interactions. Speech to text is indicated by the Automatic Speech Recognition (ASR) to Natural Language Unit (NLU) relationship. Text to speech is represented by the Natural Language Generation (NLG) to Text to Speech (TTS) components.

The ***NVTT-Shadow Simulator*** provides the game content including the payload operator (PO) user interface (UI) and the air vehicle (AV) simulation. The Simulator contains proxies for interfacing to other NVTT-Shadow architectural components.

One Semi-Automated Forces (***OneSAF***) provides the contextual elements of the game including the terrain and simulated entities on the terrain. These entities are observed by the payload operator playing the game. OneSAF includes the Night Vision Image Generator (NVIG).

Computer communications between components is through the Distributed Interactive Simulation (DIS) network.

The ***Measurement*** Platform provides an assessment of student performance using data from the other system components as well as measures developed specifically for the NVTT-Shadow . The Measurement Platform is based on three Aptima products: Performance Dashboard, PM Engine™, and ASA. Integration of the Measurement Platform enhances the performance measurement and after action review (AAR) capabilities of the NVTT-Shadow training environment through the integration of relevant human performance measures within NVTT-Shadow for use by both synthetic entities and AAR modules. The Measurement Platform, shown in Figure A-2 consists of the Measurement Platform Controller, the PM Engine Pool Manager, the Performance Dashboard (UI), Performance Measurement Engine, and the A-Measure Server Application (ASA).

The Performance Dashboard generates scorecard and After Action Review displays for viewing visualizations of the performance data collected and processed by PM Engine and stored within the ASA. These displays are dynamically generated using the configuration information contained within the project configuration file. This includes information about the filters, data sources, and visualization the dashboard creator wants displayed in the UI. Configurable items include:

- List of included modules in the main tab bar along the top of the application
- Name of the dashboard and main tab bar tabs
- Data Sources available for visualization
- Filters which can act on visualizations
- Visualization definitions

The ASA is a framework for aggregating performance measurement data from disparate systems. The ASA uses the Human Performance Markup Language (HPML) to define performance measures, as well as to describe human performance from a variety of sources, including Aptima's PM Engine™ and SPOTLITE™. Custom data schemas and API extensions can also be added through ASA plugin modules.

The ASA Database is hosted on SQL Server 2008 or higher and uses Entity Framework 5.0 as the ORM tool providing the Data Model. The database can evolve during release of updates to the system without the loss of data. Web services which require user authentication are used to store and retrieve data from the ASA. Data stored in the ASA is used to drive both real-time and after action review web based dashboards.

The Performance Measurement (PM) Engine is a real-time analysis tool that can query, filter and perform calculations on any type of data. PM Engine calculates and assesses measures from data generated during simulation-based exercises. It measures, assesses, tracks, and reports performance to provide real-time feedback. The PM Engine uses HPML to handle data from a variety of sources, such as a DIS network, log files, and SPOTLITE™. PM Engine serves as the data analysis component of the measurement platform.

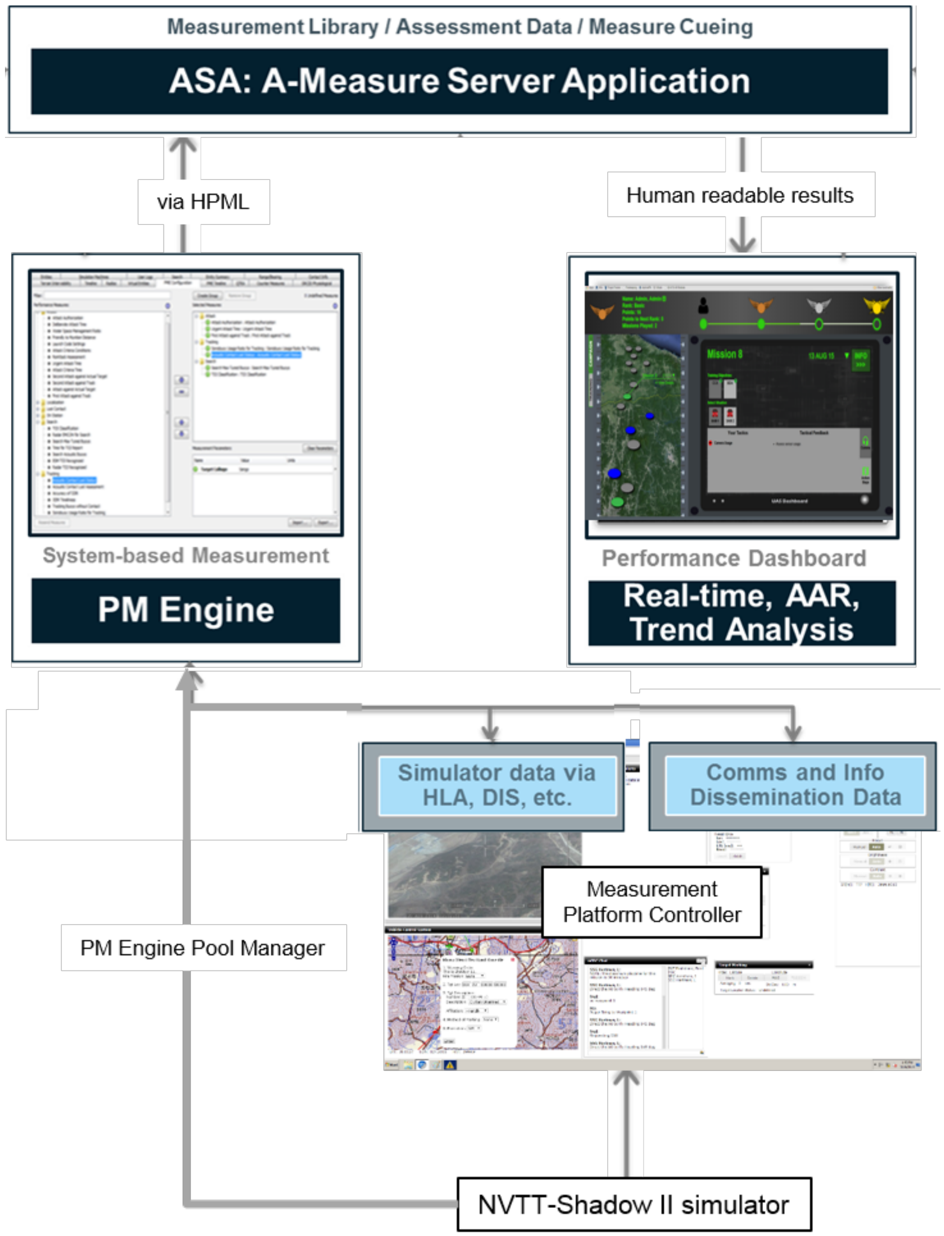

Figure A-2. NVTT-Shadow Measurement Platform with ASA, PM Engine, and the Performance Dashboard

For NVTT-Shadow, the PM Engine processes performance data consistent with SPOT, close combat attack (CCA), BDA, indirect fire (IF) and target handoff (THO) Reports and calculates corresponding performance measures. The PM Engine also assesses Communication Reports on four dimensions including Accuracy, Completeness, Order and Duration. Measures are assessed on a three tier discrete rating scale.

APPENDIX B

NVTT-SHADOW SYSTEM CURRICULUM

Aircrew Training Manual (ATM) tasks were mapped to the 10 training missions and review of missions led to alignment of campaign missions to Soldier tasks by system type (Shadow or Grey Eagle). The team created activity diagrams to describe the action between the player and the constructive entities as well as the branches and sequels in the action for: Indirect fire, close combat attack, target handover with and without the laser target marker (LTM) and laser designator range finder (LDRF) and remote Hellfire designation. In addition, for the campaign missions, the team completed modifications to the OneSAF mission scenarios to include mission scripts with recommended injects providing students with addition mission situational awareness.

The NVTT-Shadow system curriculum includes both Training Missions (Table B-1) and Campaign Missions (Table B-2). Training Missions are used to provide instruction in how to play the game. Campaign Missions include scored game play.

Table B-1
Training Mission Curriculum

	Training Mission Name	**Description/Objectives**
1	Maneuver the Air Vehicle	Description: The mission is to effectively communicate with the AO in order to change the AV heading, fly to a predetermined waypoint, fly to a Military Grid Reference System (MGRS) grid location, fly to specified Lat/Long, increase and decrease AV altitude (minimum Alt is 4000 Ft MSL) and increase and decrease AV speed. There is no hostile threat.
2	Operate Payload Sensors	Description: The mission is to effectively use the payload Electro Optical and Infra-Red sensors to observe targets on the ground. There is no hostile threat.
3	Track a Moving Target	Description: The mission is to effectively use the payload Electro Optical and Infra-Red sensors to observe moving targets. There is no hostile threat.
4	Create a Target	Description: The mission is to use the Payload Operator Video stare point to create target in the system database. There is no hostile threat.
5	Submit Reports	Description: The mission is to effectively send reports to the Brigade Operations and Intel section using standard SPOT and BDA report formats. There is no hostile threat.
6	Target Handoff with and without Laser Pointers	Description: The mission is to conduct target handover at night with the Laser Target Marker and Laser Designator. There is no hostile threat.

	Training Mission Name	Description/Objectives
7	Call for and Adjust Indirect Fire	Description: The mission is to locate targets in EA DAGGER and then call for and adjust indirect fire.
8	Call for Fire Air (Rockets and Guns/Cannon)	Description: The mission is to locate targets in EA DAGGER and then request rocket and cannon fires from the Apache element using the 5-Line format.
9	Call for Fire Air (Missile)	Description: The mission is to locate targets in EA DAGGER and then request missile fires from the Apache element that will be on station South of EA DAGGER.
10	Request Close Air Support	Description: The mission is to locate targets in EA DAGGER and then request close air support from the Eagle element that will be on station SE of EA DAGGER.

Table B-2

Campaign Mission Curriculum

	Campaign Mission Name	Description/Objectives
1	Route Reconnaissance	Description: Shadow platoon conducts a route recon of MSR CHEVY to support the movement of 1SBCT vehicle convoys from the Air Port of Entry (APOE) to their initial assembly areas. There will be traffic jams, construction avoidance issues and suspicious vehicular activity to report and to guide ground elements to for security interdiction.
2	Area Reconnaissance	Description: Shadow platoon conducts an area recon of AA STEELERS to support the 1st Bn, 1SBCT occupation of AA STEELERS. There will be suspicious vehicular activity to report and the payload operator will guide ground elements for security interdiction.
3	Area Security	Description: Shadow platoon conducts an area recon of AA RAVENS at night to support the 3rd Bn, 1SBCT occupation of AA RAVENS. There will be a different set of traffic jams, construction avoidance issues and suspicious vehicular and ground activity to report and to guide ground elements to for security interdiction.
4	Route Reconnaissance to Support NEO.	Description: Shadow will conduct a route recon of ROUTE PONTIAC to support the movement of a NEO vehicle convoy from PYONGGANG to the international boundary, continuing to the end of ROUTE PONTIAC where it terminates at the north end of MSR CHEVY. They will continue to provide overwatch

	Campaign Mission Name	Description/Objectives
		and route security/traffic avoidance information until the convoy reaches the airfield.
5	Screen International Border	Description: Shadow Platoon conducts a screen of AC REFEREE to support the Atropian border forces and provide intelligence to prevent/halt incursions across the IB.
6	Route Reconnaissance/Surveillance (IED)	Description: Shadow platoon is on station to provide route security along MSRs EDSEL and STUDEBAKER as directed by the TOC. An insurgent improvised explosive device (IED) team will either be identified by Shadow or a ground security element and Shadow will be tasked to guide an OH-58 Team to engage the insurgent IED team.
7	Route Reconnaissance/Surveillance (IED)/Support Downed Aircraft/Personnel Recovery	Description: Shadow will be on station in the vicinity of CP 24 to provide route security From CP 24 to CP 54 on MSR CHRYSLER in support of the subordinate battalions' supply convoys. During the mission the TOC will direct Shadow 11 to execute an on order mission of downed aircraft recovery support.
8	Suppression of Enemy Air Defense	Description: The mission is to find and destroy an SA-15 launcher platoon with indirect fires or an AH-64 Attack Team if necessary.
9	Area Reconnaissance	Description: Conduct an area recon and correctly identify and track the target convoy, and either follow it to its destination or hand over the target to another observer if the target crosses a BCT boundary.
10	Area Reconnaissance	Description: In coordination with the CAB Conduct an area recon of TAI 3 with the intent of remotely designing Hellfire missiles that will destroy a convoy while it is halted for refueling in SEPO.
11	Area Reconnaissance	Description: Conduct an area reconnaissance of TAI 4 in order to identify new threat convoy routes and locate convoy refueling points and rest halts on the convoy route.
12	Support Attack	Description: Support a USSOCOM attack to eliminate the Arianan nuclear weapon threat and remove the warheads for technical evaluation. Search for and report threat air defense assets with the capability to influence operations at the airfield. On order observe a TAI to provide early warning of Arianan forces reacting to the attack.
13	Area Reconnaissance	Description: Conduct an area reconnaissance of EA LEAVENWORTH to locate and destroy SRBMs which have

	Campaign Mission Name	Description/Objectives
		departed their missile storage areas and are deployed to field hide locations.
14	Area Reconnaissance	Description: Conduct an area reconnaissance of EA LEAVENWORTH to locate and destroy SRBMs while they are moving from hide location to the field location. The payload operator will receive information from the TOC as it becomes available to help narrow the search area within EA LEAVENWORTH.

APPENDIX C

DESIGN FOR TRAINEE ENGAGEMENT

The NVTT-Shadow game has been designed using principles of gamification with the goal to implement a user interface in the system that creates intrinsic motivation for the student trainee Payload Operator (PO). To this end, a Mission Map feature was developed to illustrate the game-like journey of a student from each completed campaign mission to the next. The illustration uses a satellite image of the Area of Operation and shows the missions ahead. A game control interface was designed that leverages the Mission Map as the primary navigational structure for choosing game missions and feedback.

The Mission Map feature was implemented to depict the game-like journey of a student from each completed campaign mission to the next. The depiction uses a satellite image of the AO and shows the missions ahead. Knowing one's progress in relation to the larger "story" serves to provide helpful motivation for voluntary game players. A new game control interface was designed with the Campaign Mission Map as the primary navigational structure. Additionally, game play was improved with significant improvements to the joystick functionality.

- Using a split interface with the Mission Map and an information Dashboard sharing the screen, the user can click mission nodes on the Mission Map and view information about that mission instantly in the Dashboard.
- Within completed missions, the Dashboard displays an intuitive hierarchical structure which allows the user to select a desired training objective (e.g., Target Handover) and then individually select each target of that training objective. Details for each target contain three types of feedback for Communication measures and Action Steps (behavioral measures):
  (a) at-a-glance stoplight ratings on key dimensions of performance,
  (b) narrative developmental feedback, and
  (c) for communication analysis, an audio comparison of the user's report to a sample expert's report.
- When future missions (i.e., those not yet completed) are selected on the Mission Map, the Dashboard displays a preview description of the mission and related training objectives, along with a "Play Now" button for starting the mission directly from this control panel. Thus, the new integration between the Mission Map and an information Dashboard enhance the game-like feel of the simulator, and provide consistent motivation to complete all missions.

Scoring strategies were also designed to support gamification and student engagement:

- Award students points for more and better performance
- Advance student game rank through points
- Award badges for virtual reward
- Advance rank minimally for non-failing mission play

- Require increasing amounts of expert performances for higher ranks

The NVTT-Shadow supports game control in the following ways:

- Joystick control of the simulated UAS
- In-game control via a Voice Interface
- Audio input and audio feedback supported using a microphone equipped headset

Figure C-1 illustrates a typical student (PO) workflow.

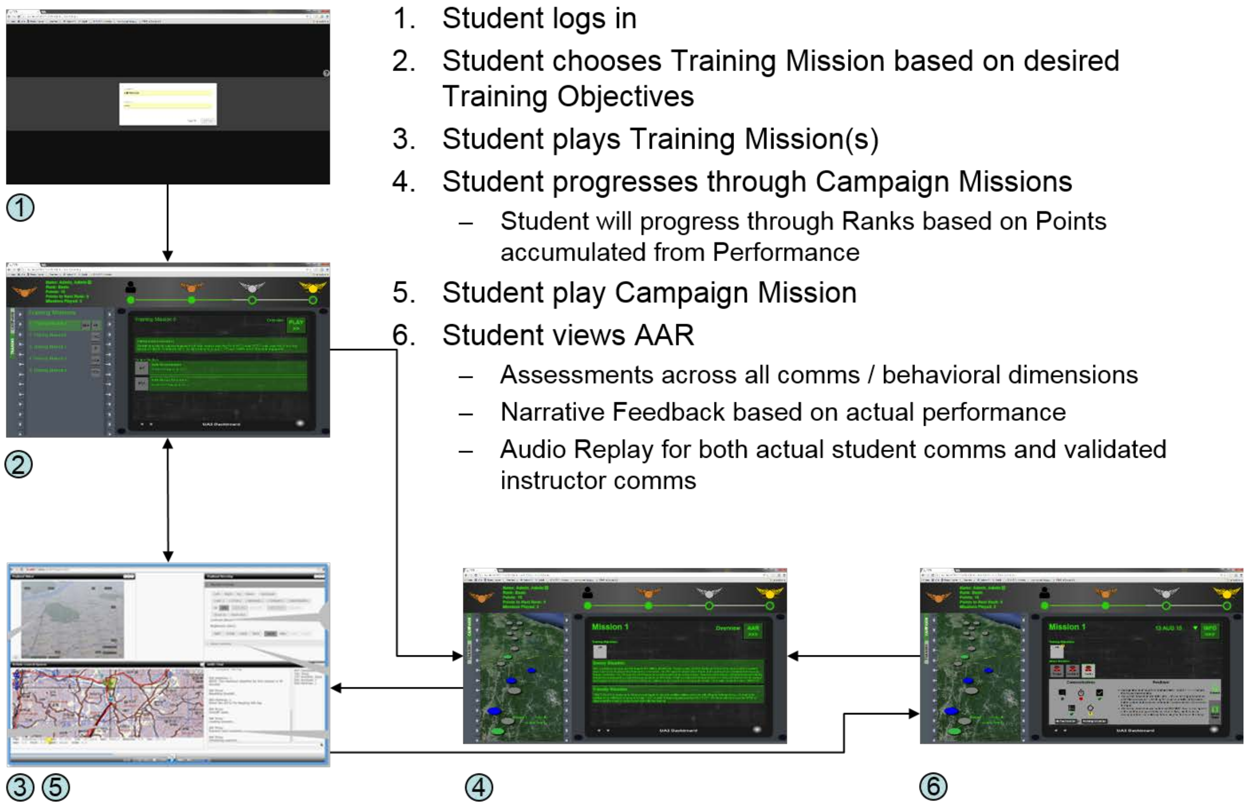

Figure C-1. Student Workflow

The UI, shown in Figure C-2, has a similar layout to typical UAS Ground Control Stations (GCS). The layout includes the payload (sensor) view in the upper left hand corner. Directly below the payload view is the map or Compressed ARC Digitized Raster Graphics (CADRG) view. The right hand side of the display includes various payload operator controls, feedback, and chat.

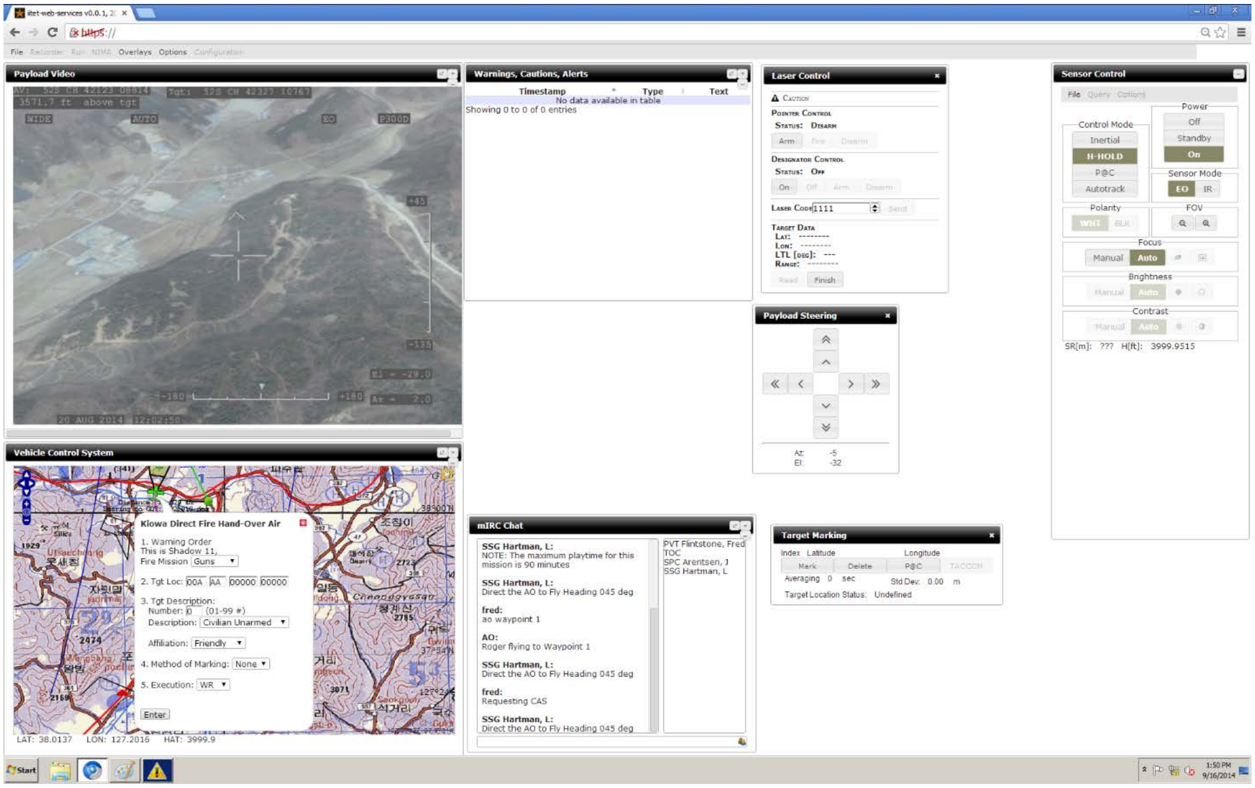

Figure C-2. NVTT-Shadow UI

The NVTT-Shadow game interface includes engaging graphics for scores and feedback, as well as a status bar showing progress toward next rank. Rank is portrayed using a color-coded insignia signifying levels of achievement including Unranked, Bronze, Silver, Gold, and Platinum. To follow the theme of the system, the insignia was chosen to resemble aviation wings. The concept is to create intrinsic motivation within the student with game elements such as being able to achieve rank promotion with high performance.

The game control interface (NVTT-Shadow UI) provides access to the set of Training Missions and set of Campaign Missions via tabs along the left side of the display.

The Campaign Missions tab uses the Campaign Mission Map as the primary navigational structure. The Mission Map feature illustrates the game-like journey of a student from each completed campaign mission to the next. The illustration uses a satellite image of the Area of Operations (AO) and shows the missions ahead. When future missions (i.e., those not yet completed by the student) are selected on the Mission Map, the Dashboard will display a preview description of the mission and related training objectives, along with a “Play” button for starting the mission directly from this control. The integration between the Mission Map and an information Dashboard will enhance the game-like feel of the simulator, and provide consistent motivation to complete all Campaign missions.

The Training Missions tab provides a list of training mission titles, a list of training objectives rehearsed in each training mission, completion status, and the ability to play or replay the training mission. Unlike Campaign Missions, Training Missions do not occur in a continuous narrative and thus do not require a map-based scorecard for the user interface.

Figure C-3 illustrates a display for which the user has selected Training Mission 8.

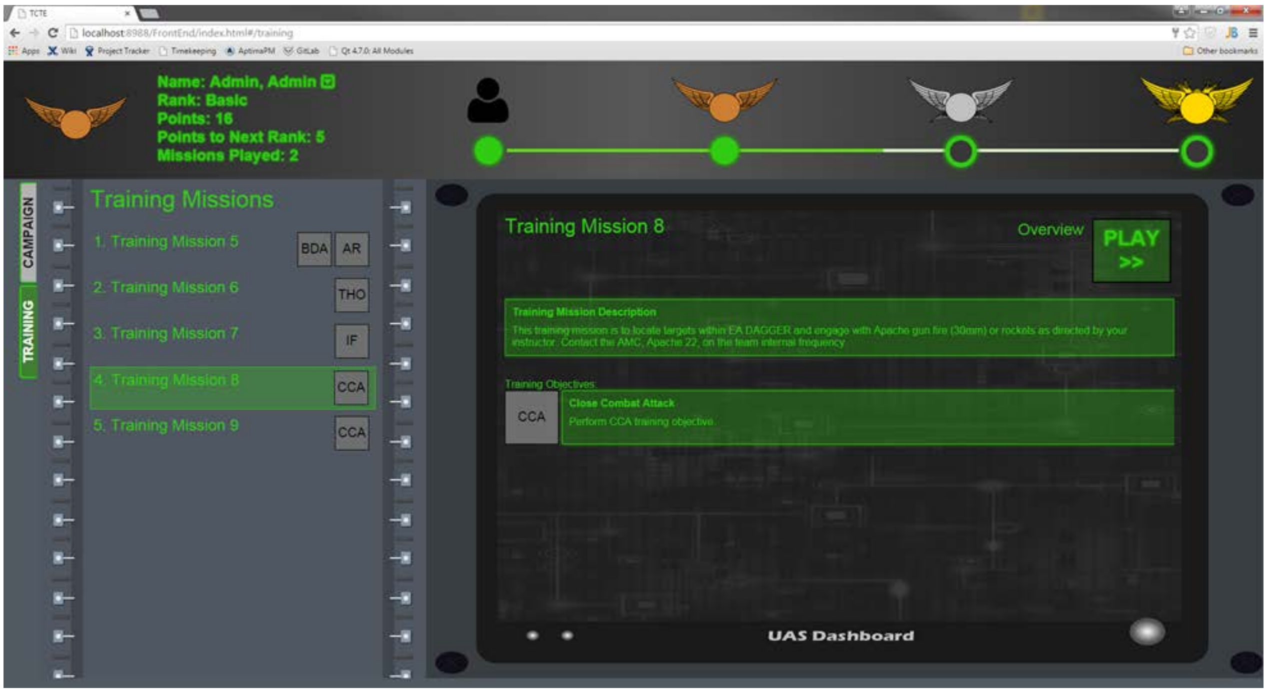

Figure C-3. Training Mission 8 Display

Using a split interface (as shown in Figure C-4) with the Mission Map and an information Dashboard sharing the screen, the user can click mission nodes on the Mission Map and view information about that mission instantly in the Dashboard. Within completed missions, the Dashboard displays an intuitive hierarchical structure which allows the user to select a desired training objective (e.g., Target Handover) and then individually select each target of that training objective.

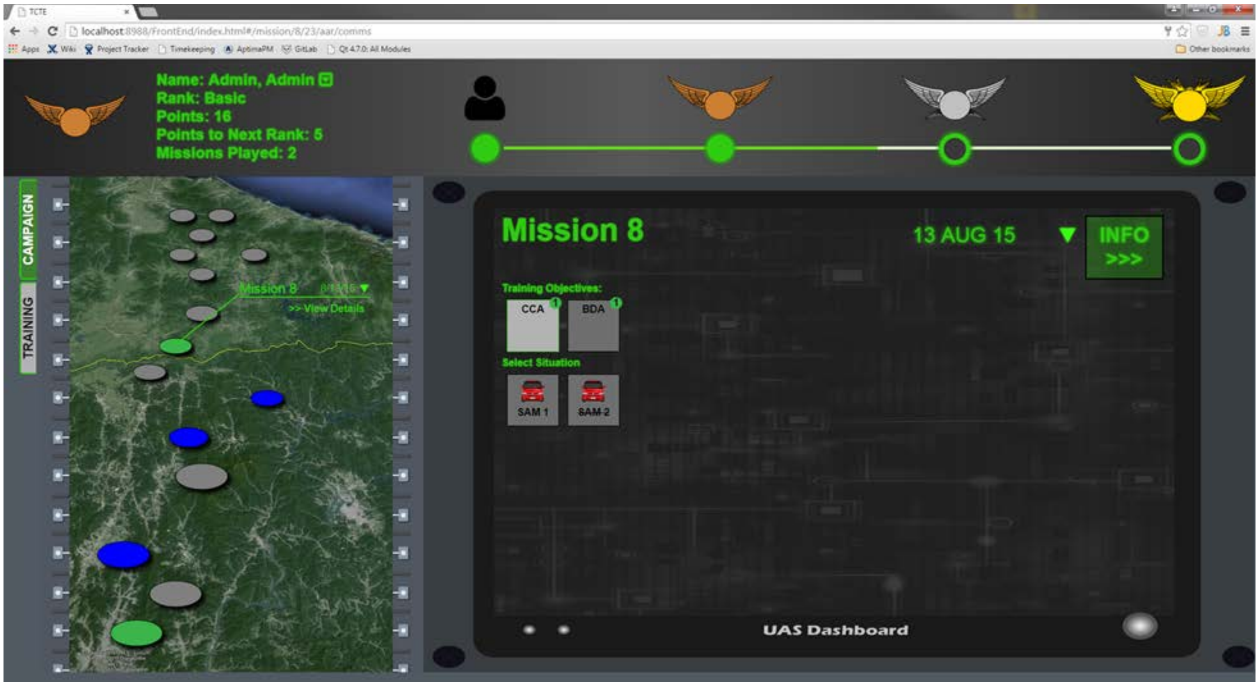

Figure C-4. Split Interface with Information Dashboard and Mission Map

Details for each target contain three types of feedback for Communication measures (Figure C-5) and Action Steps, or behavioral measures (Figure C-6).

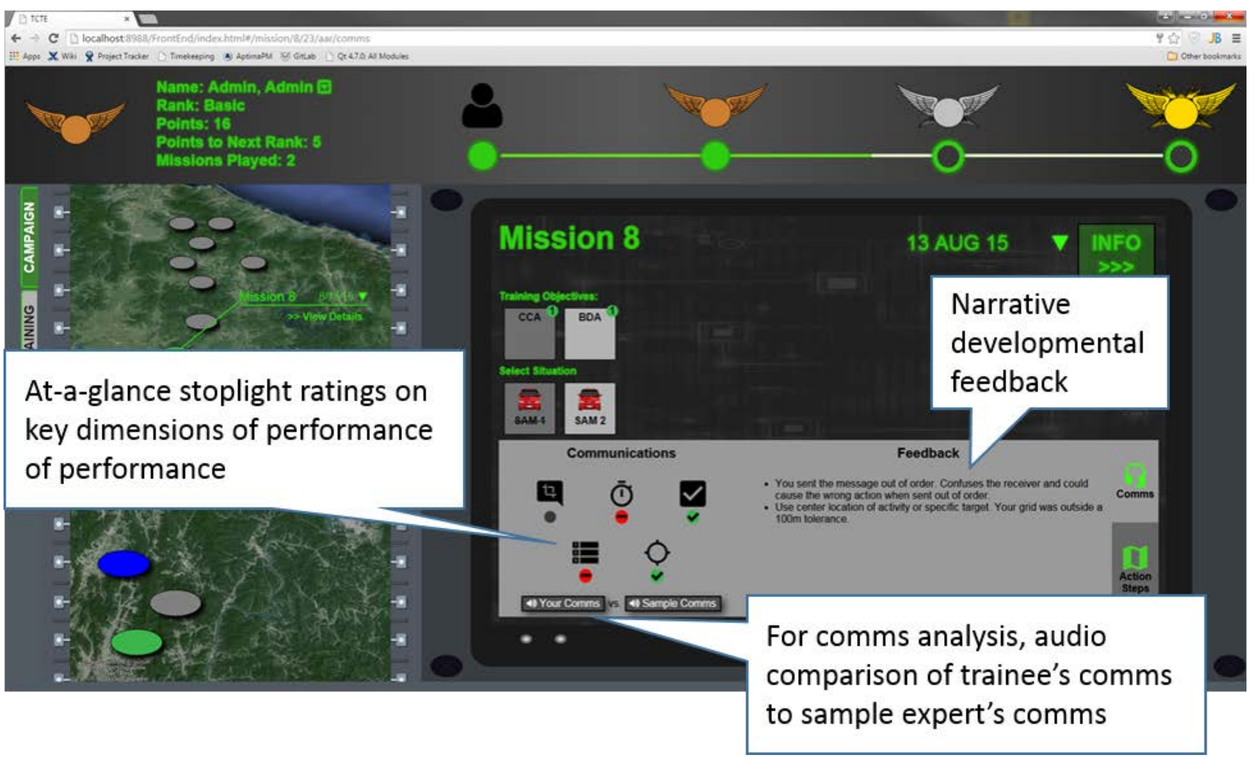

Figure C-5. Communications Measures Feedback

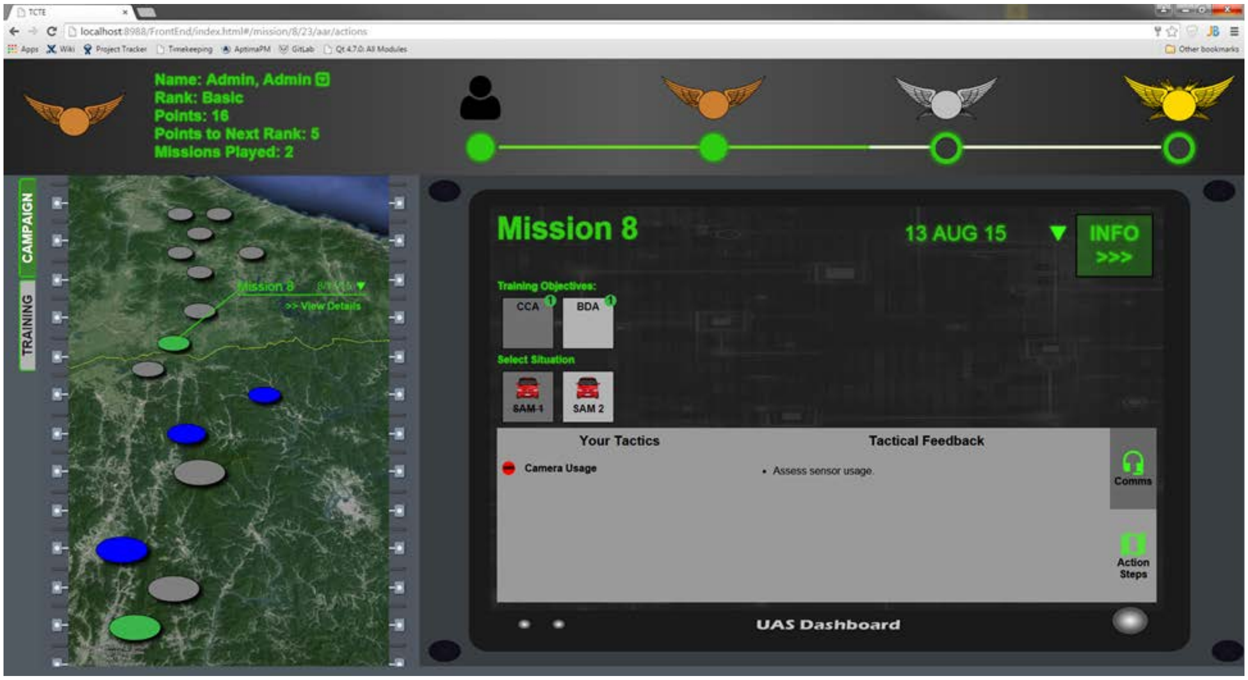

Figure C-6. Behavior Feedback

As shown in Figure C-7, an After Action Report (AAR) is available for review following a training session.

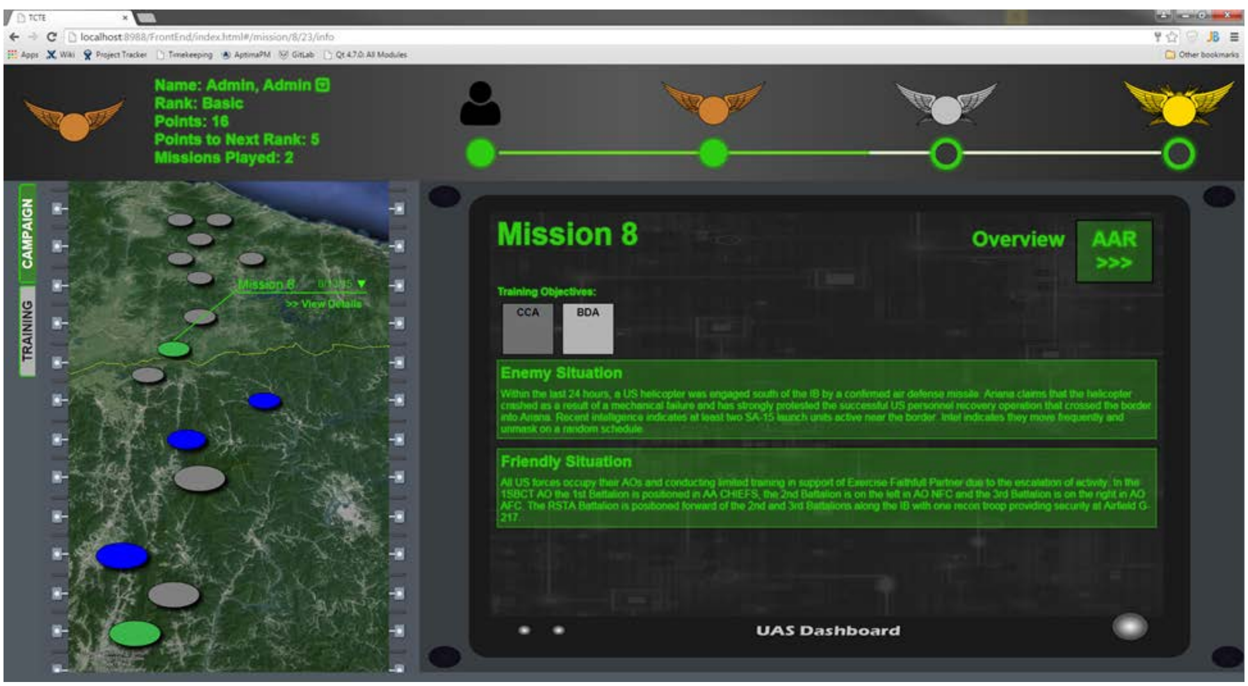

Figure C-7. After Action Report (AAR)

APPENDIX D

MEASURES DEVELOPMENT – COMMUNICATIONS VS. BEHAVIOR

Measures

A measures analysis was conducted which identified measures to be supported and their corresponding simulation data dependencies. There are two types of measures used to calculate scores: communication and behavioral. Communication is measured along four dimensions:

- Accuracy: Were the report specific items reported accurately? Accuracy of communication encompasses both the content and form of a piece of trainee communication. That is, system must decide: (1) whether a trainee's utterance expresses the content that is required at a given point in the scenario, and (2) whether the utterance meets the relevant protocols for military communication in this environment. Accuracy measures use an alignment-based approach to provide a general and robust solution to highly variable voice input. Words from trainee utterances are aligned to answer templates generated by SMEs with corresponding "goodness" scores. In the case of lack of matches because of trainee use of different lexical items or phrases (i.e. noisy trainee input), semantic resources for natural language processing will be used to search for semantically similar words and phrases, and the semantic similarity of these words/phrases will be used in the accuracy score.
- Completeness: Were all report items given? This is measured principally by whether the utterance fills all appropriate slots in the expected utterance template.
- Order: Were the report items given in the correct sequence? Evaluating order is based on protocol for the sequencing of particular dialog acts. This measure draws on the spoken interaction history of the interaction manager and the NLU module's classification of utterances into expected dialog acts. The communications analysis module computes how closely the ordering of information fits the optimal, prescribed ordering (weighting the evaluation of performance by distance to the protocol-based ordering, if desired).
- Duration: How long did it take the student to give the report? Duration is assessed by considering how the student payload operator made proper use of the time that they had to communicate the message, and if it was communicated at the right time within the context of the mission. It is important to note here that duration does not equate to speed, since it is important that the student understand the urgency related to each communication and makes good use of the time afforded to them so that they can form accurate and complete communications.

Communications measures development included:

- refining a reference corpus from SME input to support accuracy and completeness measurement;
- implementing a new algorithm for matching trainee utterances to a reference corpus to support accuracy scoring;
- refining rule-based accuracy scoring algorithm to better reflect important features of utterances; and
- implementing a PM Engine plugin for communications measures with support for SPOT report analysis.

Behavioral measures focus on the appropriate coupling of communications to trainee actions in the simulation. Behavioral (“action step”) MUM-T measures were defined for Campaign Mission 8 and the data to be used in the calculation of those measures were estimated, making the measures ready for development. The team developed corresponding narrative feedback for these measures that is tailored for the user, with lower scores receiving additional explanation and developmental recommendations for improvement.

The accuracy scoring approach was revised to account for SME input, with changes focused on scoring target descriptions relative to a hierarchy of specificity of description (e.g., “vehicle” is less specific than “sedan” and is scored lower) and scoring activities in SPOT reports based on a wider variety of criteria (neutral/colloquial, specificity).

APPENDIX E

GAME SCORING

A system was implemented for aligning and scoring utterances from trainee reports and providing feedback to the trainee. Great care was employed to design the system to be extensible and tailorable by utilizing human readable and editable tables of information to govern each step in the process. This allows for relatively painless incorporation of new campaign missions in the future or tailoring the current system to most effectively train Soldiers as trainers deem necessary.

Scoring Mechanics

A scoring framework was developed for "binning" measure scores, then aggregating into scores for training objectives, missions, and game-related points earned (toward "rank" promotion).
The following are further details about how scoring is accomplished in NVTT-Shadow .

For each performance measure (i.e., result), measure triggers (i.e., when a measure should be calculated), measure components (e.g., objects, attributes), and calculations to be performed on components to produce the result must be defined. Measures can be binary (pass/fail) or stoplight (excellent/acceptable/deficient).

A scoring strategy is defined for each measure as follows. Every measure gets a green/yellow/red assessment (bin), for which performance thresholds (e.g., high/excellence = >X, med/acceptable = X, low/deficient/needing improvement = <X) are identified. The thresholds for each measure are defined in the measurement templates, tailored for missions, as needed. Aggregating green/yellow/red measure scores at each level requires setting a "passing" threshold (e.g., what percent of possible score is passing?), and determining whether weights (priority) should be applied to any scores.

The use of Stoplight scoring at each level requires a measure to be defined as either pass/fail (Green/Red) or Green/Yellow/Red. For stoplight scoring, each Training Objective (TO) must be pass/fail (Green/Red) and every Mission is pass/fail only. Rank is advanced minimally for non-failing mission play. Increasing amounts of expert performance over time are required for higher ranks.

Determining the scoring strategy for TOs involves an aggregation of Behavioral and Communications measures. Measure bin scores are added together. The highest possible score is calculated, given the number of measures (and any weights) selected for that TO. If a particular measure is clearly higher-stakes than the others for that mission, weighting may be applied. These cases are rare. A pass/fail (green/red) threshold is defined for each TO template (i.e., what percent of possible score is a "passing" %), and a pass/fail score is calculated for each TO attempt based on that threshold.

The scoring strategy for Missions (aggregation of TOs) involves adding together the various TO scores for that mission. Weighting is only used if the TO is new to the student - that's

when the TO matters most (best opportunity for developmental feedback). A pass/fail threshold is defined for each Mission (or across missions) using one of the following options.

- Option 1: Certain number of TO attempts passed for this mission (e.g., requires 4 of 6 TOs to pass this mission)
- Option 2: What percent of TO points is required for "passing" a mission? (i.e., similar percentage-style calculation with threshold as used in TO scoring)

Additionally, the system applies a strategy for earning Career/Rank points (+ points for good performance, - points for bad performance). These points are used to determine Rank promotion criteria.

Feedback Strategy

At the conclusion of each mission, the trainee receives tailored feedback in several forms based on the performance measure results.

- Stoplight indicators: Green, red (and yellow, when applicable) indicators of progress on specific missions, training objectives, and measures.
- Narrative feedback: canned text generated from patterns of scores within a mission/training objective; feedback contains description of performance plus recommendations for doing better. The feedback contains (a) 1-2 sentences on what the expected competencies were for this report; (b) short list of bullets of *how* the trainee's report was deficient; (c) instructions to listen to the Sample Report to see how that could have been reported better.
- Audio Report Comparisons: listening to Soldier's own report vs. an expert sample report; provides comparisons for cadence, emphasis, and clarity.

Utterance Scoring

An integral part of the NVTT-Shadow is the ability to access utterances of the trainee. This project has built a scoring mechanism that relies on the notion of *slot alignment* to align subsequences of words in the utterances of the trainee with exemplars of legal utterances based on each set of simulated test scenarios. The employed algorithm is a subset of the approach developed by Sultan (2014). A classification of each of the trainee's utterances is then used to score the utterances and generate appropriate feedback for training purposes.

NVTT-Shadow allows trainees to practice formation of proper communications for a variety of structured reporting types including SPOT, Battle Damage Assessment, Remote Hellfire, Call for Fire, Target Handover, and Close Air Support. This paper focuses on the assessment of the trainee's utterances and automatic generation of constructive feedback. Each reporting type is composed of structured utterances which can be decomposed into phrases, or slots. In some cases, the phrases can be further decomposed into smaller semantic units, e.g. a number or count followed by a description of targets. Since we know the details of each simulated scenario a priori, we know what the student ought to be reporting. That is, we are afforded a very restricted domain of appropriate utterances from the trainee. This makes it possible to enumerate a set of legal, exemplar phrases and semantic units for which the student

should strive. Furthermore, we can classify the legal phrases into subsets where some phrases are more desirable than others. Using this partitioning of phrases, we can apply a scoring mechanism that compares the phrases of the trainee's utterance to the subsets of legal phrases, allowing for some phrases of the utterance to be preferred over others and generating a proportional score. The sets of legal phrases can be classified in such a way that each subset of phrases has a similar deficiency, if a deficiency exists. This classification scheme can be exploited to generate constructive feedback when deficiencies in the trainee's utterance are encountered, e.g. the trainee should not refer to the target as "a group of guys", but instead use more precise and informative language like "five individuals".

Construction of Phrases

Each simulated campaign mission is composed of events. The trainee is supposed to observe the events and respond with the appropriate utterance corresponding to the correct type of report(s) for the event. The information in the utterances has a structured form. For example, a well formed utterance for the SPOT report includes descriptive observations along with the position and time of the observations. Table E-1 shows the information present in a well formed utterance for each of the report types. We refer to each of these elements of the utterance as a *slot*.

Table E-1
Slots of Utterances for Each Reporting Type

Report Type	Slots of Utterance
SPOT	Target description with count, location, time
Battle Damage Assessment	Target location, time, target description with count, status
Remote Hellfire	Warning order, *friendly location*, target location, target description with count, *remarks*
Call for Fire	Warning order, target description with count, *fire control*, *engagement*, location, time
Target Handover	Warning order, target location, target description with count, *remarks*
Close Air Support	IP/BP, heading, distance, target location, target description with count, *friendly location*, egress

Each utterance's slot in Table E-1 contains specific kinds of information that should have a reasonably structured form. Some slots have more degrees of freedom in their expression than others. For example, the *target location* has a precise, standardized format which relays information about the coordinates of the target; whereas *target description with counts* can be formulated and uttered in many equally correct ways, and therefore requires a more sophisticated natural language processing technique for assessment. This discussion concentrates on the more sophisticated and less standardized slots where the trainee has more freedom in their formulation of the utterance.

Subject Matter Experts (SME) supplied common, anticipated slot phrases that a trainee may utter under the circumstances presented in each of the simulated events. The SME groups the phrases into classes where some classes of phrases are more desirable than others, i.e. the trainee should receive a higher score for uttering phrases in some classes than in other suboptimal classes. One of the classes should be reserved for optimal phrases, i.e. there should

be a class of phrases that the trainee should be striving for and receives the maximum score for uttering a phrase from this class. Each of the other classes should be grouped such that the phrases in that class have a common deficiency that make those phrases suboptimal, e.g. the trainee used less formal or descriptive language to describe the target such as “a group of guys” instead of “five individuals”. Additionally, the classification system can be applied at any level of resolution that is needed to sufficiently express the ranking of phrases. In other words, this system affords the definition of as many classes as are necessary to properly rank the phrases under the simulated event to which is pertains. The classification system is flexible and extensible.

Note that the *target description with count* slot is common to all of the report types. As the name suggests, this can be decomposed into sub-phrases. For example, the utterance, “five individuals carrying RPGs,” can be decomposed into a count of “five,” a target description of “individuals,” and a description of the activity they are engaged in, “carrying RPGs.” Decomposing the *target description with count* slot in this way allows the enumeration of phrases by the SME to be simplified by enumerating the legal count/number phrases, target description phrases, and activities phrases separately, reducing the slot to phrases that can be more easily enumerated independently. Table E-2 shows a fabricated example of this decomposition for illustrative purposes.

Table E-2
Example of Abbreviated Legal SME Phrases for SPOT Report

Numbers		Descriptions		Activities	
Description	**Class**	**Description**	**Class**	**Description**	**Class**
one	1	persons	1	carrying rpgs	1
two	1	personnel	1	holding rpgs	1
three	1	individuals	1	carrying weapons	2
four	1	dismounts	1	looking suspicious	3
five	1	men	2		
six	1	Soldiers	2		
seven	1	troops	2		
eight	1	insurgents	2		
nine	1	guys	3		
single	1	folks	3		
several	2				
lot of	2				
bunch of	2				
some	2				
many	2				

The SME’s lists of common phrases with their classifications are encoded in a spreadsheet. The NVTT-Shadow simulation software is designed to directly read the spreadsheet. This allows the SME and trainers to work with a human readable format that the simulation software can interpret to employ the rest of the scoring algorithm based on the encoded information. Most importantly, this allows alterations or additions to easily be made to

the legal phrases and classification system at any time, including after system delivery via text file updates.

Slot Alignment

Each event in the simulator has corresponding report types that should be uttered by the trainee. Each event's report type has an associated list of slot phrases as defined by the SME. The goal of slot alignment is to find the SME's slot phrase that most closely fits the trainee's actual utterance. This is accomplished through text alignment.

Generally speaking, text alignment is the task of identifying textual segments in different sources of text that have similar semantic meaning. It is an important topic to many fields within natural language processing including automatic machine translation, information retrieval, question answering, and many others. Text alignment is an active area of research which has attracted increased attention recently. Relevant portions of the method proposed by Sultan *et al* 2014 paper were employed to address the alignment of the trainee's spoken phrases of the trainee with the set of the SME's legal phrases.

Text alignment is applied to a pair of phrases from a slot where one of the phrases is the trainee's utterance for that slot and the other is an entry in the SME's list of legal phrases. There is a preliminary step performed before the main alignment algorithm is applied. The location of stop words is identified in both phrases. Stop-words are common words in a language that often hold little semantic value and are often present to inform the reader/listener of grammatical structure, e.g. *the*, *a*, *of*, etc.. Other words, not in the stop-word list, are considered to be content words. These content words are assumed to hold the more semantic meaning, and are therefore more productive for the purpose of aligning phrases.

The core of the text alignment algorithm is designed to find the longest continual subsequences, i.e. *n*-grams, of the content words in the shorter phrase that map to subsequences in the longer phrase. Since longer continual subsequences are preferred, the algorithm searches for matching *n*-grams where *n* is initially equal to the length of the shorter phrase, then *n* is reduced until either there are no more matches, *n* equals zero, or all content words have been accounted for. In this way, a set of content alignments is produced between the two phrases.

An alignment scoring algorithm is used to score each pair of aligned phrases. The scoring algorithm, which is the Sorensen-Dice Similarity Metric, is shown below:

$$Score_{alignment} = \frac{2\,|C|}{|A| + |B|}$$

Where $|A|$ is the number of content words in one phrase, $|B|$ is the number of content words in the other phrase, and $|C|$ is the number of words that were aligned. Note that $|C| = |A \cap B|$ when there does not exist duplicate words in either phrase.

A text alignment score of the trainee's phrase is computed for each of the phrases in the SME's phrase list for the applicable slot. The SME's phrase with the highest score is selected as the best aligned phrase for the slot and is used in for utterance scoring and feedback generation.

Utterance Scoring

There is a *weight* associated with each of the SME's entries in Table E-3. The weight numerically encodes the correctness or acceptability of the phrases being uttered in a given slot and is used in the scoring algorithm. A higher weight means that the phrase is more correct than lower weights. Often, these weights are direct functions of the SME's assigned class for the phrase, but the code is implemented in such a way that this does not necessarily need to be the case. That is, if it becomes convenient or necessary to customize the weights in the future, the code is written to facilitate the changes to the spreadsheet.

Table E-3
Example abbreviated SME descriptions for SPOT report with weight and constraint

Numbers

Description	Class	Weight
one	1	1.000
two	1	1.000
three	1	1.000
four	1	1.000
five	1	1.000
six	1	1.000
seven	1	1.000
eight	1	1.000
nine	1	1.000
single	1	1.000
several	2	0.500
lot of	2	0.500
bunch of	2	0.500
some	2	0.500
many	2	0.500

Descriptions

Description	Class	Weight	Constraint
persons	1	1.000	5
personnel	1	1.000	5
individuals	1	1.000	5
dismounts	1	1.000	5
men	2	0.666	5
Soldiers	2	0.666	5
troops	2	0.666	5
insurgents	2	0.666	5
guys	3	0.333	5
folks	3	0.333	5

Activities

Description	Class	Weight
carrying rpgs	1	1.000
holding rpgs	1	1.000
carrying weapons	2	0.666
looking suspicious	3	0.333

An additional piece of information is encoded in the slot's description column. Since each event within a campaign mission reflects a specific simulated scenario which portrays a specific number of targets that the trainee is reporting on, we encode the correct number of targets for each description type. That is, if the description was *persons* and the simulated event has five people depicted, we encode that there are five people. We call the correct number of targets the *constraint*. This offers a way for the scoring and feedback code to compare the trainee's numeric description to the actual number of specific types of targets in the simulated event. If the trainee says there are more *persons*, for example, than exist in the simulation, the score and feedback ought to reflect this mistake. Also, the number of targets can be independently defined for individual types of targets. So, for example, we can encode an event

where there are five people and two tanks by constraining the people descriptions to five and the tank descriptions to two.

Table E-3 shows examples of the weights and constraints. Note that Table E-3 is the same as Table E-2, but with the additional weight and constraint columns.

Scores are separately produced for two types of phrases. If the trainee uttered "*five persons carrying RPGs*", then the first scores phrases that are aligned to a number or count and aligned to a target, e.g. "*five persons*" where "*five*" was aligned to the number and "*persons*" was aligned to the target description. The second type of phrase is the aligned activity that the targets are performing, e.g. "*carrying RPGs*".

The score for the count and target is a function of the weights associated with the number and target description, the presence of the number and target, and the numeric constraint. The utterance of a correct cardinal number is directly rewarded, as specifying the definite number of targets in the simulator should be a goal of the trainee. The score for the count and target is shown below:

$$Score_{target} = (\delta_{card} + \delta_{cont} + k_{descr} \cdot w_{descr}) \ / \ k_{norm}$$

Where δ_{card} and δ_{cont} are indicator functions for the presence of a cardinal number and the case where the numeric constraint is not violated, respectively. w_{descr} is the weight from Table 5 for the description, and k_{descr} is a coefficient that determines the importance of the weight. k_{norm} is a normalizing constant so that the score is linearly scaled between zero and one. In our case, $k_{descr} = 2$, and $k_{norm} = 1 + 1 + k_{descr} = 4$. The values of the coefficients can be changed as desired to assign more weighting to the description element.

The score for the activity portion is simply the associated weight assigned in Table E-3. That is:

$$Score_{activity} = w_{activity}$$

If multiple target or activity phrases are present in a trainee's utterance, the scores above for each are averaged to produce a combined score. That is:

$$Score_{multiple} = \frac{1}{N} \cdot \sum_{i=0}^{N} Score_i$$

These target and activity scoring methods are applied to all the appropriate report slots present in Table 3.

Feedback Generation

The software is able to generate feedback on the trainee's utterances that is specific to each event, report type, and slot. In the previous sections, the trainee's utterances were decomposed and aligned with phrases in the SME's legal phrase lists. Each phrase in the list had an associated class. As explained in the *Construction of Phrases* section, the classes can be used to group phrases together such that each class of phrases has common deficiencies. Messages can be defined such that when a phrase with a certain deficiency is encountered, as defined by its class, an appropriate feedback message is displayed to the trainee informing them of the

deficiency. The contents of the feedback message can be defined in concert with the definition of the classes. This affords the ability to alter or tailor the feedback messages in the future by forming new classes and constructing appropriate messages that pertain to the simulated event that the trainee is practicing.

In addition to using the classes to provide feedback, other types of feedback can be generated. For example, in the number and target portion of a slot, the numeric constraint can be used to generate feedback if the trainee mentioned more targets than exist in the simulated event. Also, the absence of phrases can trigger feedback informing the trainee of the missing information. In other words, if either δ_{card} or δ_{cont} are equal to zero during the utterance scoring (as defined in the previous section), feedback can be generated to inform the student that they should use cardinal numbers to specify the number of targets or that they have specified more targets than exist in the simulated event.

Table E-4

Feedback Table which Specifies Feedback Messages According to Features and Classes of Phrases

Size Feedback			Activity Feedback		
Feature	**Class**	**Feedback**	**Feature**	**Class**	**Feedback**
NoNumber	0	The number of target persons was missing.	NoActivityClass	0	A valid activity description was missing.
NoDescription	0	A description of the target persons were missing.	ActivityClass	1	N/A
UsedCardinalNumeral	2	Your description should be specific about the number of target persons.	ActivityClass	2	The correct activity is "carrying an rpg 7". Be more specific.
UsedCardinalNumeral	1	N/A	ActivityClass	3	Your description of the activity of the enemy was too vague.
ExceededDescrNumConstraint	1	You mentioned too many targets.			
ExceededDescrNumConstraint	2	You mentioned too many targets.			
ExceededDescrNumConstraint	3	You mentioned too many targets.			
DescriptionClass	1	N/A			
DescriptionClass	2	The correct description was persons, people, or individuals. Specific descriptions of persons involved in the “activity” provide better context from			

		which to make decisions.
DescriptionClass	3	The correct description was persons, people, or individuals. You should avoid using informal language. Specific descriptions of persons involved in the "activity" provide better context from which to make decisions.

The feedback generation mechanism is implemented by producing a list of *features* while a slot is being analyzed in the utterance scoring section of the code. Features essentially indicate properties of a phrase that may elicit feedback, such as the trainee exceeded the numeric constraint or a portion of the phrase is missing. Combining the features with the classes of aligned phrases provides a structured way to select appropriate and constructive feedback to the trainee. Table E-4 shows an example of features, classes, and the corresponding feedback message, and is intended to work with the information defined in Table E-3. Note that in Table E-4, an artificial class of zero was introduced to indicate the absence of that information from the utterance. Also, feedback messages of N/A indicate that there was no deficiency with this feature and therefore, no constructive feedback is necessary.

As an example, if the trainee uttered, "*five persons carrying RPGs*", the phrase would be decomposed into a target count of "five", a target description of "persons", and an activity of "carrying RPGs". The trainee has uttered all relevant portions, the number is the correct cardinal number according to the *constraint*, and each portion belongs to class 1 which in this case corresponds to generating no feedback. But if the trainee instead uttered, "Guys looking suspicious", they are missing the number of targets and they used a class 3 target description and activity. This would trigger a *NoNumber* feature, a class 3 *DescriptionClass* feature, and a class 3 *ActivityClass*, which in turn generates the feedback, "The number of target persons was missing," "The correct description was persons, people, or individuals. You should avoid using informal language. Specific descriptions of persons involved in the *activity* provide better context from which to make decisions," and "Your description of the activity of the enemy was too vague", respectively